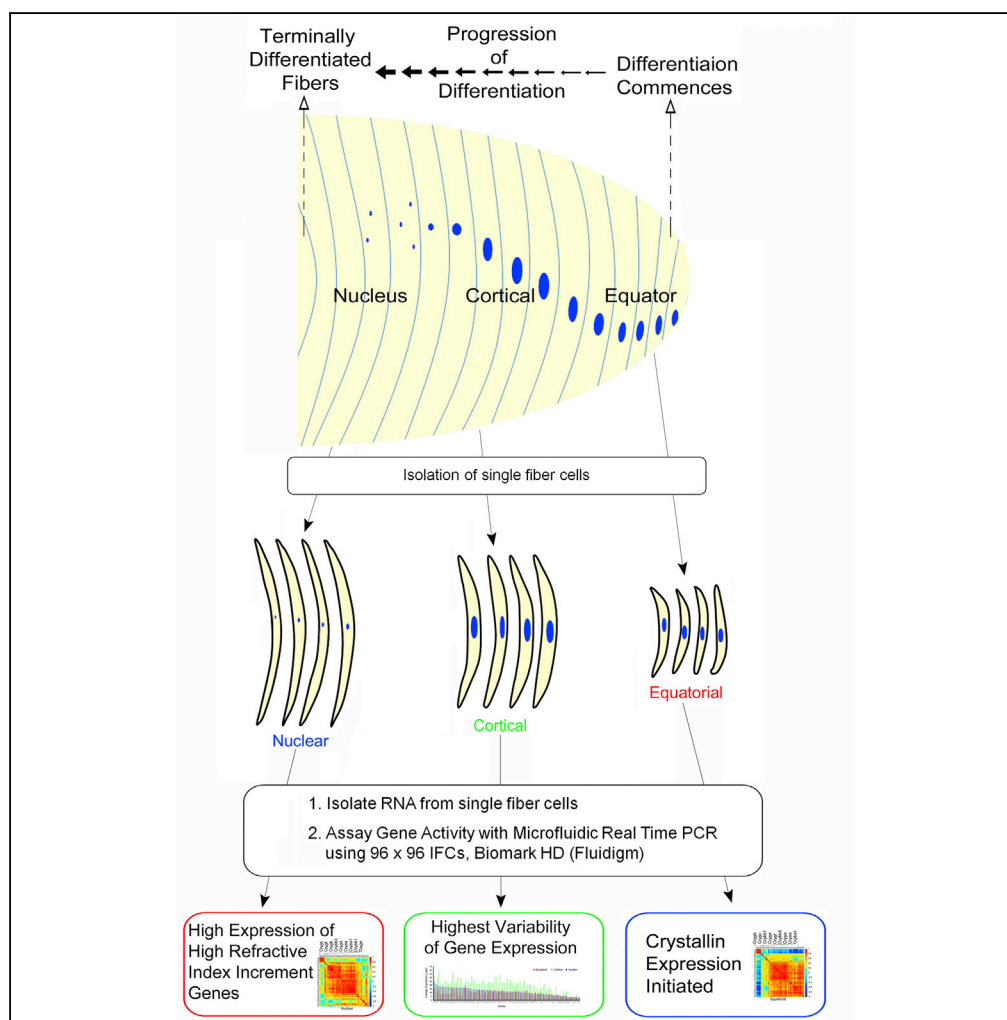**Article**

Spatial Analysis of Single Fiber Cells of the Developing Ocular Lens Reveals Regulated Heterogeneity of Gene Expression

Rajendra K. Gangalum, Dongjae Kim, Raj K. Kashyap, Serghei Mangul, Xinkai Zhou, David Elashoff, Suraj P. Bhat

sbhat@mednet.ucla.edu

HIGHLIGHTS

Transcriptional heterogeneity attends epithelial cell differentiation to fiber cells

Early crystallin expression suggests determinism in terminal fiber differentiation

Genes for high-refractive-index-increment proteins are coordinately expressed

Molecular heterogeneity makes every fiber cell a unique refractive unit

iScience

Article

Spatial Analysis of Single Fiber Cells of the Developing Ocular Lens Reveals Regulated Heterogeneity of Gene Expression

Rajendra K. Gangalum,[1] Dongjae Kim,[1] Raj K. Kashyap,[1] Serghei Mangul,[2] Xinkai Zhou,[3] David Elashoff,[3] and Suraj P. Bhat[1,4,5,6,*]

SUMMARY

The developing eye lens presents an exceptional paradigm for spatial transcriptomics. It is composed of highly organized long, slender transparent fiber cells, which differentiate from the edges of the anterior epithelium of the lens (equator), attended by high expression of crystallins, which generates transparency. Every fiber cell, therefore, is an optical unit whose refractive properties derive from its gene activity. Here, we probe this tangible relationship between the gene activity and the phenotype by studying the expression of all known 17 crystallins and 77 other non-crystallin genes in single fiber cells isolated from three states/regions of differentiation, allowing us to follow molecular progression at the single-cell level. The data demonstrate highly variable gene activity in cortical fibers, interposed between the nascent and the terminally differentiated fiber cell transcription. These data suggest that the so-called stochastic, highly heterogeneous gene activity is a regulated intermediate in the realization of a functional phenotype.

INTRODUCTION

There is extraordinary heterogeneity in gene expression in single cells; whether it is noise or consequential activity remains unclear (Symmons and Raj, 2016). It is essential to understand the biological consequences of transcriptional heterogeneity at the single-cell level (Snijder and Pelkmans, 2011), but the very isolation of a cell without its neighbors (Klein et al., 2015; Macosko et al., 2015) brings into question the status of the cell as a relevant unit that contributes to the phenotype of a tissue. It is, however, important to recognize that single-cell transcriptomics has allowed the identification of previously unknown functional states and phenotypes (Glotzbach et al., 2011) (Klein et al., 2015; Macosko et al., 2015; Wucherpfennig and Cartwright, 2016; Yamanaka et al., 2013). At the current state of our knowledge the utility of this molecular cataloging of individual cells or cell types, nonetheless, remains unrealized because of the absence of a concurrent morphological, developmental, and/or spatial context.

A rational link between transcriptional heterogeneity and tissue phenotype therefore remains tentative (Lein et al., 2017). A serious shortcoming has been the general lack of correspondence between the single-cell phenotype and the functional organization of the tissue. Keeping in view all the technical limitations of preserving the status of a single cell as a component of a multicellular, functional entity (tissue), we present the developing eye lens as a paradigm for spatial transcriptomics.

Contrary to the perception of an inanimate bag filled with protein, the ocular lens is a highly organized cellular structure, which is remarkably simple in terms of its cellular composition; it contains only two cell types, the progenitor anterior epithelium and the fiber cells derived from it (Figure 2A) (Bassnett et al., 2011; Cvekl and Zhang, 2017). The epithelium differentiates at its edges (lens equator, Figure 2A), adding new fiber cells on an already existing fiber mass; its youngest fiber cells are thus at the periphery, and the oldest (terminally differentiated), at the center or in the nucleus, which makes the visual axis of the lens. In the adult lens, 95% of the volume of this tissue is occupied by fiber cells (Bassnett et al., 2011; Costello et al., 2016), which synthesize crystallins to achieve transparency and maintain a specific refractive index (RI) (Iribarren, 2015).

A critical feature of this phenotype (transparency), which is set up very early in development (Peetermans et al., 1987) is that it is characterized by a gradient of RI, which follows the gradient of protein concentration, low in the periphery and highest in the center/visual axis (Campbell and Hughes, 1981; Chakraborty et al.,

[1]Stein Eye Institute, Geffen School of Medicine, University of California, Los Angeles, CA 90095-7000, USA

[2]Department of Computer Science and Human Genetics, University of California, Los Angeles, CA 90095-7000, USA

[3]Department of Medicine, University of California, Los Angeles, CA 90095-7000, USA

[4]Brain Research Institute, University of California, Los Angeles, CA 90095-7000, USA

[5]Molecular Biology Institute, University of California, Los Angeles, CA 90095-7000, USA

[6]Lead Contact

*Correspondence: sbhat@mednet.ucla.edu

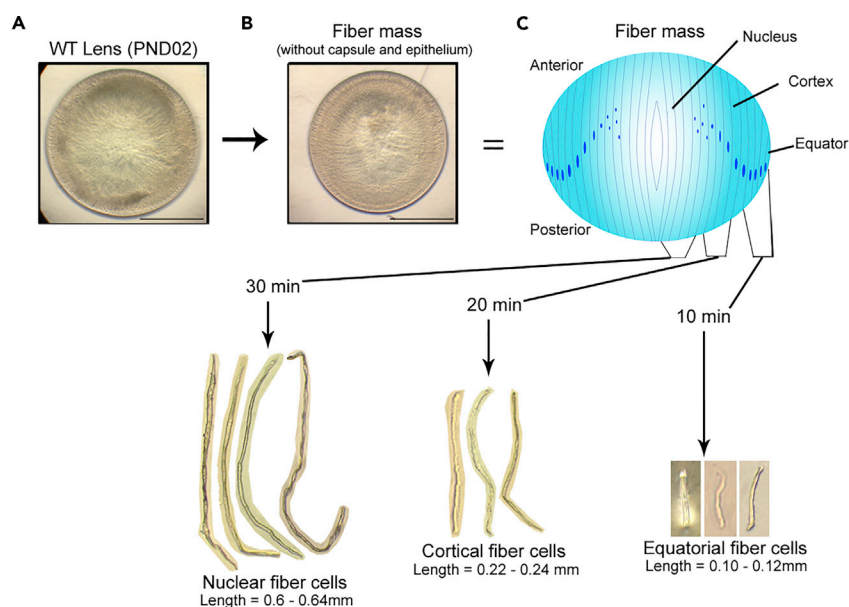https://doi.org/10.1016/j.isci. 2018.11.024

Figure 1. Isolation of Single Fiber Cells from Two-Day Old Mouse Lens

(A) A picture of the whole postnatal day two (PND02) mouse lens.

(B and C) (B) Lens (fiber mass), without the epithelium and the capsule. (C) The schematic (blue) depicts three distinct regions (outer equatorial, middle cortex, and inner lens nucleus). The fiber mass was processed to collect individual fiber cells by incubating it with shaking (Transparent Methods). The outer equatorial fiber cells (length: 100–120 μm) were collected in first 10 min; the cortical fiber cells (length: 220–240 μm) were collected in next 10 min, and finally the nuclear fiber cells (length: 600–640 μm) were collected in the last 10 min. Scale bar, 500 μm. In this study, we did not see a direct correlation between fiber cell length and gene activity with respect to 94 genes (including crystallins).

2014; Philipson, 1969; Pierscionek et al., 1987) (Pierscionek and Regini, 2012). A smoother gradient of RI assures pinpoint focus without spherical aberration. Importantly, therefore, every fiber cell becomes an optical unit that must reconcile to its spatial status and molecular uniqueness, attained through specific crystallin expression (Mahendiran et al., 2014) to refract light into the eye to make vision possible.

Given this developmental history, the lens lends itself to the isolation and study of single fiber cells, region by region in a spatially contextual fashion (Figures 1 and 2A). For example, the gene activities in single fiber cells can be followed spatially from the equator to the cortex to the nucleus (center) of the lens; the center of the lens has the highest protein concentration and the highest RI as its phenotype. At postnatal day 2 (PND02), differentiation from the overlaying epithelium has already generated an equatorial region (nascent fibers), a cortical region (differentiating fibers), and a nuclear region (terminally differentiated fibers), which makes the future visual axis. Importantly, the lens at this stage is pliable and therefore allows us to peel off fiber cells, one at a time, by gentle shaking of the fiber mass. The younger superficial fiber cells come off quicker, whereas the deeper fibers take longer. Based on the time it takes for fiber cells to come off, we have cataloged them as equatorial fibers, cortical fibers, and nuclear/central fibers (see Transparent Methods section).

In this investigation, we have studied single fiber cells of the PND02 mouse lens (Figure 1) using microfluidic quantitative qRT-PCR in a Biomark microfluidics system (Fluidigm, Inc.). The analytical capability and low technical noise of this system allows decisive assessment of transcript levels. We have focused on known crystallin genes (all 17 of them) and other seventy seven genes of known relevance to the lens phenotype (Table S1).

RESULTS

There Is Unexpected Heterogeneity of Gene Expression in Individual Fiber Cells in the PND02 Lens

Single fiber cells (n = 446), isolated from PND02 mouse lenses were interrogated with probes for 94 genes. Unsupervised clustering of the data reveals remarkable heterogeneity (Figure 2B). As expected, the
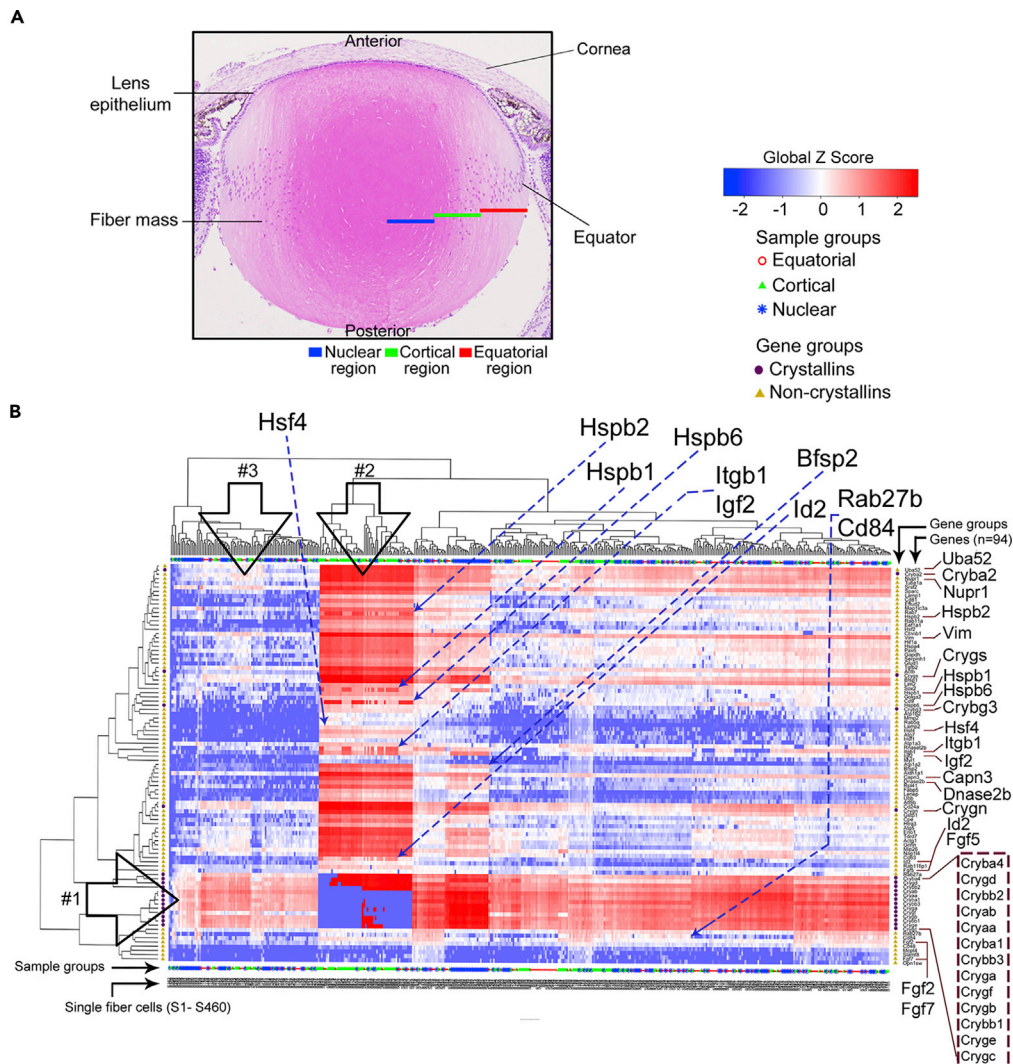
Figure 2. Unsupervised Clustering Reveals Pervasive Heterogeneity in the Ocular Lens Fiber Cells

(A) A cross section of the postnatal day two (PND02) mouse lens. We developed a procedure based on temporal release of fiber cells from the lens that allowed us to collect individual fibers (Figure 1) and assign them to one of the specific states/stages of the developing lens (indicated by colored bars; each bar, 200 μm; red, equatorial; green, cortical; blue, nuclear (see Transparent Methods in the Supplemental Information).

(B) Heatmap generated from the gene expression data (446 single fiber cells and 94 genes including 17 crystallins and 77 non-crystallins). Notice two clusters of high expression, open arrow #1 (13 crystallins) and open arrow #2 (58 fiber cells). The arrow #2 clusters fiber cells, which show very low expression of crystallins but high expression of non-crystallins; these fibers are predominantly cortical (green, follow sample groups on the x axis). Note that *Hsf4*, the major heat shock transcription factor of the postnatal lens is predominantly expressed in these fibers (indicated by a dotted arrow). Note also that in the left side of the figure (open arrow # 3, n = 93), the fibers show very poor expression of non-crystallins. Single fiber cells are annotated at the bottom of the heatmap in alphanumeric values (S1–S460); 388 single fiber cells (87%) express all 17 crystallin genes; ~10 genes (including *Mmp2, Rab5a, Lamp2, Atp1a3, Myl1, Atp1a2, Mcpt4, Fgf7, Slamf8, and Opn1sw*) show significantly low expression in all fiber cells. Long broken arrows with gene names (*Hsf4, Hspb1, Hspb2, Hspb6, Itgb1, Igf2, Id2, Bfsp2, Rab27b, and Cd84*) indicate examples of heterogeneity. Predominant expression of a gene in a specific group of fiber cells gives those fiber cells their molecular identities. Other genes of interest are indicated in bigger fonts on the right. The sample group colors, red (equatorial), green (cortical), and blue (nuclear), provide spatial information. See also Figures S1 and S2.

prevalent expression of crystallins is unmistakable (open arrow #1). A cluster of 13 crystallins is seen (Figure 2B, dotted rectangle); two other crystallins seem to associate with other genes, for example, *Cryba2* with *Uba52* and *Crygn* with *Cd24a* (Figure 2B). Similar associations involving same genes are also seen in region-specific clustering analyses (Figures S2A–S2C); they are not altered by subtle lens-to-lens (animal to animal) variations (Figure S1). Euclidean distances can change based on scaling and average expression levels, yet we see consistency in these gene associations, in clustering done with 446 fiber cells (Figure 2B) and in clustering done with about 150 each, region-specific fiber cells (Figures S2A–S2C). Another important observation about the data in Figure 2B is that the cluster of 13 crystallins is retained in the equatorial (Figure S2A) as well as in the cortical fiber cells (Figure S2B), but it is reduced to a cluster of nine in the nuclear fibers (Figure S2C; see also Figures 7A–7C).

Among the non-crystallins, remarkably, *Uba52*, a hybrid gene with an ubiquitin 5′ domain and a ribosomal *Rpl40* domain on the C terminus (Kobayashi et al., 2016), is one of the highly expressed genes (Figure 2B); comparatively, ubiquitin B is very poorly expressed. *Fgf2* is expressed in almost all fiber cells. This is significant because Fgf2 is known to be intimately associated with lens morphogenesis (McAvoy and Chamberlain, 1989). *Fgf7* (Weng et al., 1997), on the other hand, is generally poorly expressed (here it is expressed in about 30% of the fiber cells). Interestingly, Rab27b (an important member of Rab GTPase family, involved in exosome secretion [Ostrowski et al., 2010]) and Cd84, a surface glycoprotein (a member of the signaling lymphocyte activation molecules [SLAMs] and used as a differentiation marker on hematopoietic progenitor cells (Zaiss et al., 2003), show elevated expression only in specific group of fiber cells (Figure 2B, dotted arrow). A number of other interesting non-crystallin gene activities are indicated in Figure 2 in a bigger font (on the right side of the figure).

What, however, commands attention is the absence of high crystallin expression within a group of fiber cells, which predominantly come from the cortical (green) region (Figure 2B, open arrow #2). These cells make 13% (n = 58) of all fiber cells analyzed. Interestingly, this group of cells expresses a large number of non-crystallin genes, at a much higher level, as if to compensate for the lack of robust crystallin expression (Figure 2B, open arrow #2). Regional clustering of the equatorial, cortical, and nuclear fiber cells further confirms that these fiber cells are predominantly from the cortical region (Figures S2A–S2C, black arrowheads with line brackets, at the bottom of the heatmaps). It is important to note that this group of fiber cells is made of cortical fibers from all five lenses used in this study (see Figure S1). There are many gene activities, which are present only in a few fiber cells, as indicated by dotted arrows with gene names on them (Figure 2B), suggesting the presence of many "cell types," but lack of crystallin expression in lens fiber cells is unknown. Fiber cells in this group express some genes highly and others very poorly, generating fiber cells with individual identities. For example, we may simply assign a molecular tag to a fiber cell based on the expression of *Hspb1*, or that of *Hspb2* or *Hspb6* (Figure 2B). It is interesting to note that both calpain 3 (*Capn3*) (De Maria et al., 2009) and *DNase 2b* (Nishimoto et al., 2003), genes known to be expressed in deeper layers of fiber cells are highly expressed in these cells. Interestingly, these fiber cells do not express crystallins including Crygf and Cryge (De Maria and Bassnett, 2015).

Cortical Fiber Cells Show Highest Variability of Gene Expression

To ascertain the variability of gene expression in the three contiguous regions of differentiation (nascent fiber cells, cortical fiber cells, and terminally differentiated nuclear fiber cells), we calculated the variance of crystallin and non-crystallin genes (Figures 3A and 3B). The gene activity for both the crystallins and non-crystallins shows highest variability in fiber cells of the cortical region (Figures 3A and 3B, green bars). Obviously, the group of 58 fiber cells (referred to above) contributes to the variability substantially.

Interestingly, this variability contributes to variability in the fiber cell populations as demonstrated by the violin plots (probabilistic population distributions of gene activity) shown in Figure 3C. The violin plots for all the 446 fiber cells reveal bimodal or multimodal distributions in cortical fibers (Figure 3C, green, one of the columns of these distributions is indicated with an asterisk on the top) suggesting high variability and differential expression of the same gene in different sub-populations (Shalek et al., 2013). There is lower heterogeneity in the equatorial (red) and nuclear (blue) fiber cells, which are mostly unimodal (for violin and boxplots of fiber cell gene activity in individual regions see Figures S3–S5).

Principal component analysis (PCA) reveals that the three-dimensional space occupied by the expression levels of 94 genes sifts the 446 fiber cells into roughly five clusters (Figure 4A). Cluster #1 contains 58 fibers,
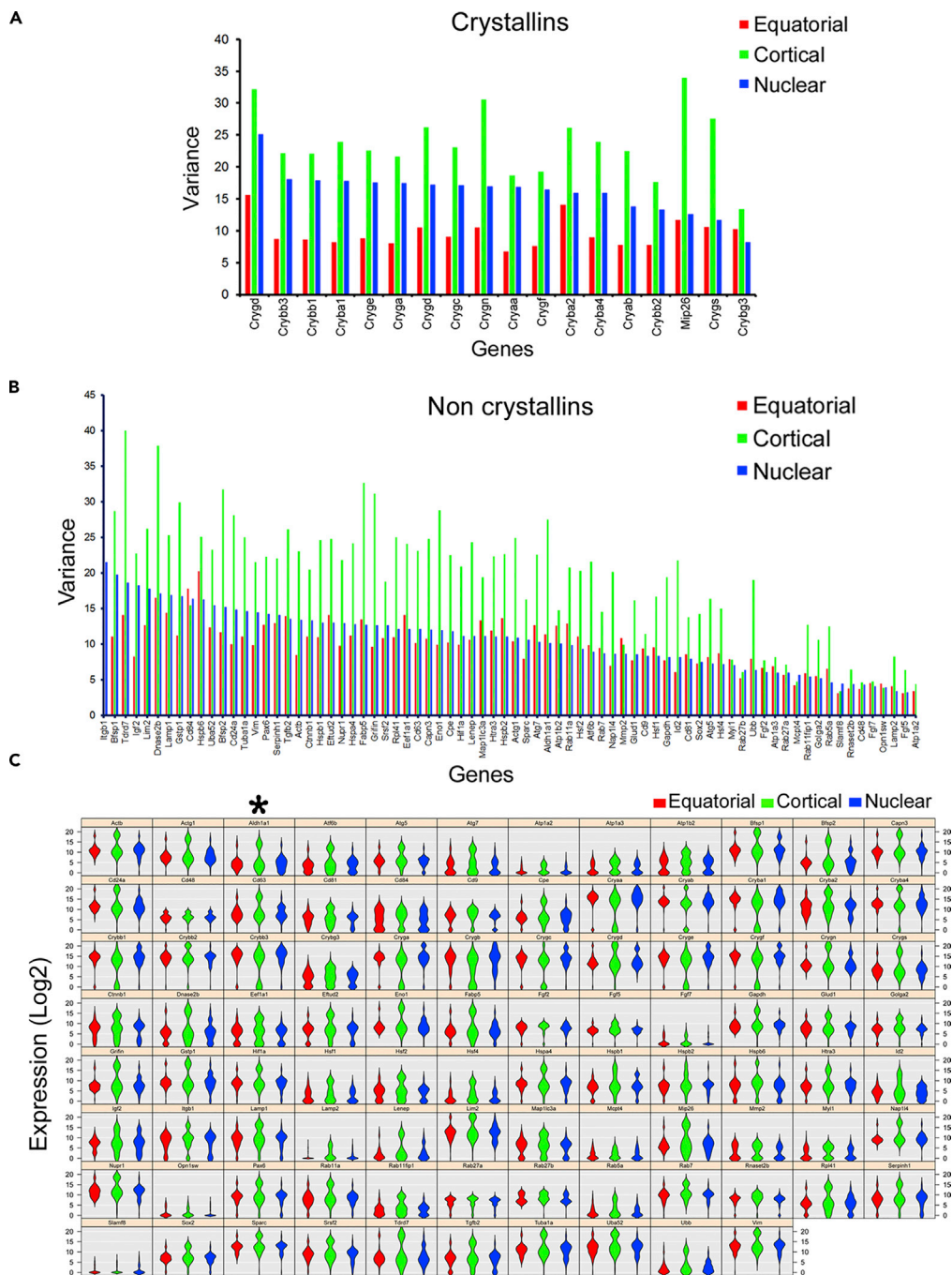
Figure 3. Gene Expression in the Developing Cortical Fibers Is Highly Variable

(A) Variance in crystallin gene expression.

(B) Variance in non-crystallin gene expression. Cortical fiber cells (green bars) show highest variance in >95% of genes analyzed compared with equatorial (red) and nuclear fiber cells (blue). The variance values are plotted based on highest to lowest in the nuclear fiber cells.

(C) Violin plots of gene expression in individual 446 fiber cells from equatorial (red), cortical (green), and nuclear (blue) regions. The distributions in equatorial and nuclear regions are similar and mostly unimodal (y axis, expression (log2); x axis, number of fiber cells). The cortical fiber cells (green) stand out (one of these columns has been indicated with * on top). These cells show bimodal/multimodal distributions indicating expression in more than one population of cells. For region-specific violin plots, see also Figures S3–S5.
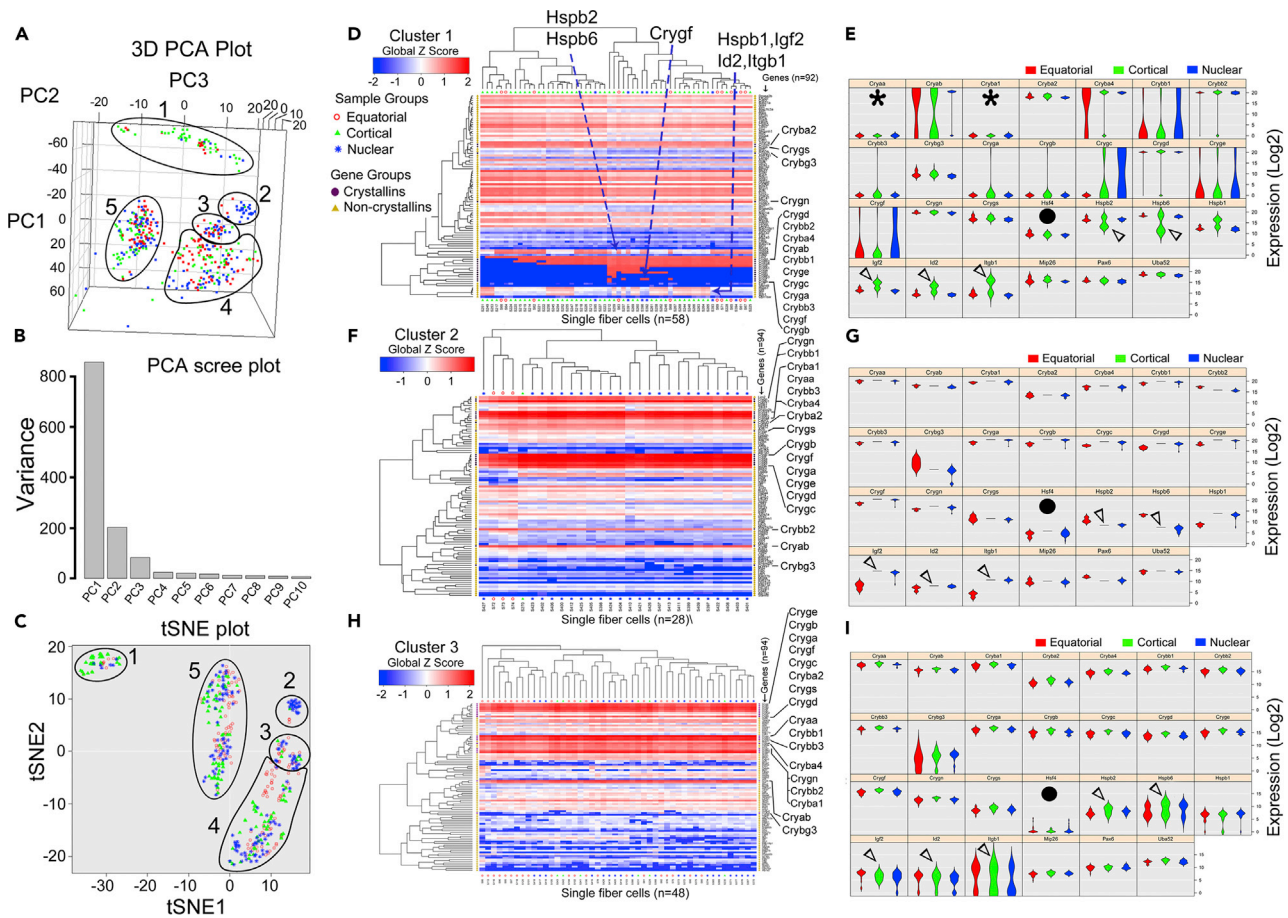
Figure 4. PCA Reveals Multiple Populations of Fiber Cells

(A) PCA compartmentalizes the expression of 94 genes in 446 fiber cells into roughly five clusters.

(B) A PCA scree plot. Cluster #1 (PC1) contributes about 86% variance; cluster # 2 (PC2), 20%; and cluster # 3 (PC3), about 8%.

(C) t-SNE (distributed stochastic neighbor embedding analysis [van der Maaten and Hinton, 2008]) scatterplot reiterates the separation of the cell populations seen in the hierarchical clustering (Figure 2B).

(D and E) (D) Heatmap and (E) violin plot (partial) for cluster #1 (n = 58, 42 cortical, 10 equatorial, and 6 nuclear). Out of 11 crystallins (D, lower right), only *Crygd*, *Crybb2*, and *Cryba4* are expressed in 44 of 58 fibers. A few of these fibers show expression of *Crygf* in cells, which express *Hspb2* and *Hspb6* (broken arrows), suggesting expression in discrete cell populations. Note also specific patterns of *Hspb1*, *Igf2*, *Id2*, and *Itgb1* expression (top right broken arrow). These fibers do not express *Cryaa* and *Cryba1* (* on violin plot). Note bimodal distribution in cortical (green) fibers (arrowheads).

(F and G) (F) Heatmap and (G) corresponding violin plots for cluster # 2 (n = 28). The majority of cells are from the nuclear (blue) region; all 17 crystallins are expressed, mostly at high levels.

(H and I) (H) Heatmap and (I) the corresponding violin plots for cluster # 3 (n = 48, 21 equatorial, 21 nuclear, 6 cortical). Note high expression of most crystallins. Examples of specific cellular distributions in these violin plots are indicated (▽) (complete violin and boxplots are presented in Figures S6–S8). Note that *Hsf4* expression is lowest in cluster #3 and highest in cluster #1(●). Clusters #4 (n = 161) and #5 (n = 140) are both constituted with equal representations from equatorial, cortical, and nuclear regions. Both these clusters show higher expression of *Cd84*, *Rab27b*, *Sparc*, and *Uba52* not shown). See also Figures S6–S8.

predominantly cortical; it contributes >80% variance. This cluster is characterized by very low crystallin expression (see Figure 4D heatmap and the corresponding partial violin plot in Figure 4E). Partial violin plots for clusters #1–3 are presented in Figures 4E, 4G, and 4I, respectively. In cluster #1, there is very little *Cryaa* (the most highly expressed crystallin in the ocular lens) and *Cryba1* (Figure 4E, asterisks), yet we see *Cryab*, *Cryba2*, *Cryba4*, *Crygd*, *Crygn,* and *Crygs,* suggesting possibly the non-refractive activities of these crystallins (Bhat, 2004).

The cluster #2 is predominantly composed of nuclear fibers, blue (n = 28), which show high crystallin expression (Figures 4F and 4G, arrowheads); it contributes 20% variance (complete violin and boxplots for clusters #1–3 are presented in Figures S6–S8; the data for other two clusters [#4 and #5] are not presented).
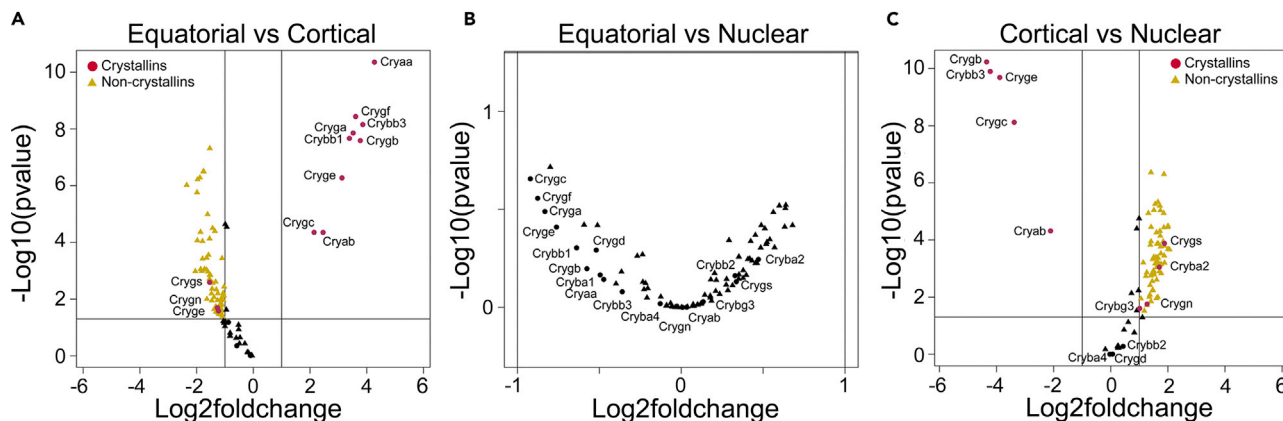
Figure 5. Differential Expression of Genes in Fiber Cells Isolated from Different Regions

(A) One-way ANOVA analysis of differential gene activity. (y axis = −log10, p values; x axis = log2 fold change). Seventy-five genes are differentially expressed between the equatorial and cortical fibers; crystallins are highly expressed in the equatorial fibers. Note that crystallins, including *Crygs, Crygn*, and *Cryge* are part of the non-crystallin gene activities (▲ not marked).

(B and C) (B) Demonstrates that equatorial and nuclear fiber cells show no significant differences in expression. Cortical versus nuclear fibers reveal that 82 genes are differentially expressed (C). Note *Crygs, Cryba2,* and *Crygn* in the cortex, along with other non-crystallin genes. ● = Crystallins, ▲ = non-crystallins.

Interestingly cluster #3 has almost equal representation from the equatorial and nuclear fibers in addition to six fiber cells that it contains from the cortical region (green). The fiber cells in this cluster show robust crystallin expression with similar violin plots, but the cluster also has cells with certain specific distributions (see open arrowheads in Figure 4I), which indicates the presence of additional cell-type variability.

The Equatorial and Nuclear Fibers Present Similar Transcriptional Profiles

A comparison of the equatorial and cortical (Figure 5A) and the cortical and nuclear fiber cells (Figure 5C) reveals a significant number of differentially expressed genes. Noticeably, we do not pick any differential activity between the equatorial and nuclear fibers (Figure 5B). It is important to recognize that cortical fiber cell gene activity follows the equatorial gene activity, both physically (side by side) as well as temporally, and nuclear activity follows the cortical activity. The equatorial and the nuclear fibers are separated by the cortical fibers physically, yet the similarity of gene activity stands out. Interestingly, it is the non-crystallin gene expression (including the expression of crystallins, like *Crygs, Cryba2,* and *Crygn*, possibly because of their catalytic roles [Bhat, 2004]) that predominates in the cortical region fibers (Figures 5A and 5C).

The data presented in Figures 6A–6C compare the average expression of each gene in the fiber cells of the three regions (the comparison is made between two regions at one time). Again, the gene activity in two separate regions, namely, the equatorial and nuclear fibers, is strikingly similar; no gene activity stands out (Figure 6B). Although the cortical gene activity is higher, the crystallins stand out in comparisons between the cortical and the equatorial fibers (Figure 6A, asterisks) as well as between the nuclear and cortical fibers (Figure 6C).

Crystallin Expression in the Nuclear Fiber Cells Is Deterministic

Hierarchical clustering (unsupervised) of all the fiber cells revealed that out of 17 crystallins, 13 crystallins made a cluster (Figure 2B, open arrow #1). The other four crystallins, namely, *Crygn, Crygb3, Crygs,* and *Cryba2*, associate with different genes (Figure 2B). Interestingly, hierarchical clustering of fiber cells in each region (Figures S2A–S2C) indicated similar patterns of clustering, which was largely maintained in the equatorial and cortical heatmaps; however, in the fiber cells from the nuclear region, the cluster of crystallins was reduced to nine (Figure S2C, dotted rectangle). Figures 7A–7C indicate that the expression of these nine crystallin genes, namely, *Cryaa, Cryba1, Crybb1, Crybb3, Cryga, Crygb, Crygc, Cryge,* and *Crygf*, is highly correlated (Pearson correlations ≥0.8–1.0). These genes first appear in the equatorial region (Figure 7A); their expression is sustained through the cortical region (Figure 7B), and they finally become part of the nuclear/central region of the lens (Figure 7C). It is noteworthy that at the single-cell level, *Cryab*
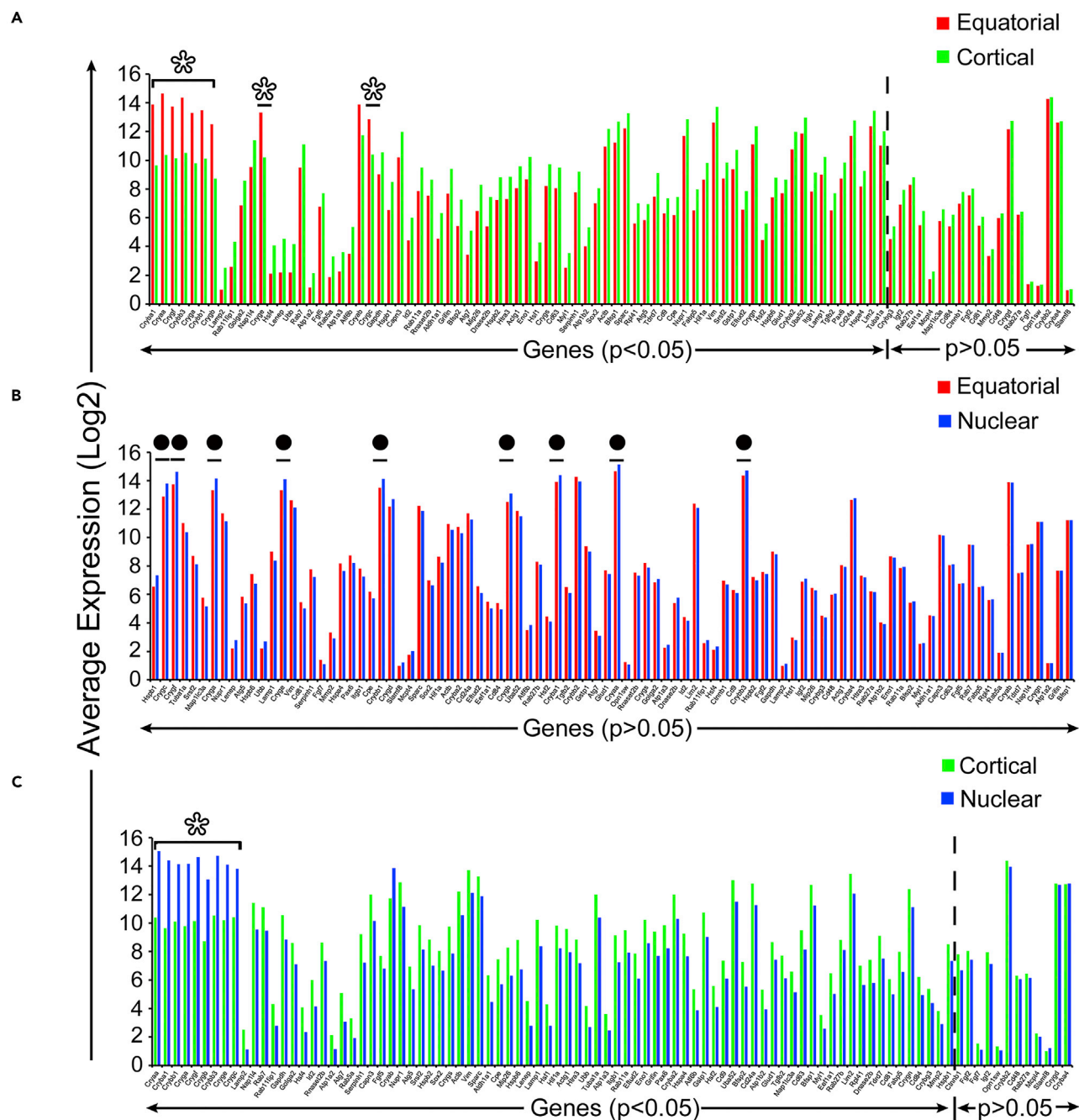
Figure 6. Equatorial and Nuclear Fiber Cells Have Similar Expression Profiles

(A–C) Comparative analyses of the expression of individual genes in fiber cells from different regions. y axis, average gene expression (log2); x axis, genes arranged based on probability (p < 0.05, significant to p > 0.05, not significant). The comparison between the equatorial and the cortical fibers (A) and the cortical and nuclear fibers (C) shows that it is the crystallins that stand out (open asterisks), although the non-crystallin activity is higher in both comparisons. (B) The expression levels in the equatorial and nuclear fibers are very similar; no gene activity stands out, including the crystallins (●).

expression is not a part of this group of nine crystallins, which further corroborates the status of the two α-crystallins (Cryaa and Cryab) as separate, independent proteins and not as a single monolithic structural protein (Gangalum et al., 2012).

Line plots of the log2 expression levels of crystallin genes and non-crystallins, across all three regions reveal that the expression of nine crystallin genes goes up in the fiber cells of the nuclear region (Figure 7D, dotted

Figure 7. Crystallin Expression in Single Fiber Cells Reflects the Phenotype of the Tissue

(A–C) Correlation matrix plots (Pearson R) for single fiber cells are presented region by region, starting from (A) equatorial to (B) cortical and finally to (C) nuclear, following the temporal and molecular progression of differentiation. The expression of nine crystallin genes (indicated on the top left in each plot) are positively correlated with each other and negatively correlated with the majority of non-crystallin gene activity in the equatorial and cortical regions (A, B, dark blue). Scale bars, +1 to −1 on the right. Similar data were obtained when all regions are analyzed together (Pearson correlations are not affected by scaling).

(D) Represents average expression (Log2) values in three regions as line plots. Nine crystallin genes, same genes as seen in the correlation plots (*Crygf, Cryaa, Crygb, Cryga, Cryge, Cryba1, Crybb1, Crybb3,* and *Crygc*) show increased expression in the nuclear/central fibers, which are part of the future visual axis, whereas other eight crystallins (*Crygd, Cryab, Cryba4, Crybb2, Crygn, Cryba2, Crygs,* and *Crybg3*) show a decrease.

(E) A schematic representation of gene activity in three spatial states of differentiation in the developing mouse lens. Note that the molecular progression starts in the equatorial fibers (red) and progresses through the cortical fiber cells (green) to finally express as terminal differentiation in the nuclear (blue) region. See also Figure S9.

rectangle) as if on cue, whereas the expression of the eight other crystallins declines along with the other 77 non-crystallins (Figure S9). It is important to recognize that the gene activity does not follow the size (length) pattern of the fiber cells (Figure 1).

DISCUSSION

In this article, we present a snapshot of the development in the mouse lens at PND02. This snapshot is not a static picture but assembled of gene activities in single fiber cells, which represent three different states/ regions of differentiation. A number of significant studies including intestinal crypts (Dalerba et al., 2011), lung epithelium (Treutlein et al., 2014), and tissues like liver (Halpern et al., 2017) and heart (Skelly et al., 2018) have all revealed developmental emergence of heterogeneity of gene expression associated with specific cell types. The data presented in this investigation, however, almost literally puts variability of gene expression, at the single-cell level, stark in the middle of the molecular progression that culminates in the functional phenotype of the tissue (Figure 3). The fiber cell morphogenesis that commences in the equatorial region ends at the center of the lens (which makes the visual axis). The data clearly suggest that the equatorial fiber cells and terminally differentiated fiber cells at the center of the lens have very similar transcriptional profiles, whereas the cortical fiber cells that separate the two show highly variable gene activity (Figures 3, 6, 7, and S9). The molecular commonality in the processes occurring at these

two morphologically separate places in the lens warrants caution because we are only looking at about 100 genes; it is possible that we do not have those genes represented here that may vary in the fiber cells of the two regions.

The very nature of the fiber cell morphogenesis in the post-embryonic lens anatomically positions the oldest differentiated fiber cells at the center and the youngest at the periphery of the lens. Thus, at one time point (namely, the PND02 lens), we can weave a pathway of molecular progression that has a temporal as well as a morphological dimension to it. As the lens keeps growing (adding fiber cells to an already existing fiber mass), a nascent fiber cell early in the development may become part of the lens nucleus (the future visual axis) in a matter of days or weeks, although this process slows considerably with age (Augusteyn, 2007). It is obvious therefore that what shapes the lens nucleus/ visual axis happens very early in development; importantly, these fiber cells and their constituents must last a lifetime, and remain transparent (Andley, 2007; Bassnett et al., 2011; Bhat, 2003; Costello et al., 2016).

What is quite interesting is that we see almost all known crystallins expressed here, but they are not uniformly expressed at the same level in all fiber cells. It is important to recognize that in the analysis of whole tissues or populations of cells, although there is no direct one-to-one correspondence seen between protein and RNA, it is becoming increasingly clearer that at the single-cell level, there is appreciable degree of correspondence between transcription and translation (Li and Biggin, 2015). If we, therefore, take the liberty of comparing the expression of crystallin genes here, with a 2D gel pattern of a newborn mouse lens (Ueda et al., 2002), an appreciation of working with the single cells versus with a population of cells becomes apparent. For example, the two proteins (Crybb2 and Crygs), which are least represented in the newborn gel pattern of the pooled lens homogenate, are highly expressed in a group of cortical fiber cells (Figure 2B, open arrow #2).

There are earlier studies wherein whole lens epithelia and whole fiber masses have been used for microarray analyses. In some studies, whole eyes were used for studying γ-crystallins (Goring et al., 1992). Studies involving *in situ* hybridizations with crystallin probes during early mouse lens development did not find many γ-crystallins in the equatorial region on postnatal day 1 (Treton et al., 1991); on the other hand, we find differential expression of γ-crystallins in individual fibers in the PND02 lens. These early pioneering investigations not only highlight the limitations of the technologies used but also point to the incongruences between population (whole tissue) studies with single-cell studies. Thus determinations of which crystallin is expressed when, within a population of cells, will only present an average picture, which may not allow meaningful functional correlations.

Previous work (Lieska et al., 1992) studiously dissected different regions of the adult bovine lens. This work demonstrated that that the *in vitro* translation of RNA isolated from the lens nucleus generates a profile similar to the protein composition of the adult lens nucleus. This clearly suggests that not only is the nuclear RNA intact and possibly functional, but more importantly, it represents the phenotype of the nuclear fiber cells. By extrapolation, therefore, the RNA transcripts we have detected in the nuclear fiber cells represent the phenotype (proteins) on these cells. Our work now provides a window into the dynamics of the differentiation process, which is harder to probe in an adult lens. The transcriptional profiles discovered in single fiber cells (Figures 5 and 6) reveal that the cellular progression from the nascent fibers to terminally differentiated fiber cells is not attended by a gradual ascension or decline of specific gene activities. Based on only 94 genes studied here, this process is book-ended by similar transcriptional profiles in the equatorial (nascent) and terminally differentiated fiber cells. The transition between the two (between the equatorial and the nuclear) is a phase of highly variable transcriptional activity in cortical fiber cells (Figure 3), suggesting that heterogeneity may be a critical intermediate in this process.

The high variability of gene expression in the cortex is not the result of gene activity variations in the same population of cells. Different populations of cells as indicated by violin plot distributions create this variability (Figures 3 and 4). It is interesting to recognize that the first cluster of cells in the PCA analysis (cluster #1, Figure 4) is predominantly composed of cortical fiber cells, which show very poor crystallin expression, and the second (cluster #2, Figure 4) is predominantly composed of nuclear fiber cells, which show high crystallin expression.

The significance of the group of fiber cells in the cortex, which do not synthesize crystallins appreciably (Figure 2B, arrow # 2, and Figure 4, cluster #1), is unknown. These cells do not express *Cryaa*, the chief crystallin of the vertebrate ocular lens (Horwitz, 2003). Proteins, other than crystallins, are known to have lower RI increments (Mahendiran et al., 2014; Zhao et al., 2011b); low crystallin expression therefore may suggest the presence of fiber cells with a slightly lower RI, which may have specific locations within the lens (Bassnett et al., 2011).

Non-crystallin gene activities, almost all of them, with very few exceptions show an increase in the cortical fiber cells (see Figure S9). These non-crystallin activities contribute to specific cellular identities (for example, see heat shock proteins Hspb 1, 2 and 6, and Igf2, Id2, Rab27b, and Cd84 in Figure 2B). This suggests a multitude of physiologies attended by molecular transformations, which prepare the fiber cells for passage to terminal differentiation. These physiologies, at one end, encompass the arrival of differentiating fibers from the equatorial region, and at the other, progression of these cells into terminal differentiation in the nucleus/center of the lens (Bassnett et al., 2011; Costello et al., 2016; Rowan et al., 2017; Subczynski et al., 2017). A critical role of cell types in region-specific appearance of pathologies (cataracts) is thus a possibility. For example, in MIP −/− (major intrinsic protein-null) mice, one of the important gene products that is downregulated is Hspb1 (Bennett et al., 2016).

We see that *Lim2* (lens intrinsic membrane protein) is a highly expressed gene in the cortical fiber cells, followed by *Cd24a, Bfsp1* (filensin), vimentin, *Uba52* (a hybrid ubiquitin-ribosomal protein Rpl40 gene), and *Capn3* (calpain 3). These gene products point to a large number of physiological activities, which encompass rearrangement of the membrane, cytoskeleton, proteolysis, and intercellular signaling. More work is required to separate and identify these activities for a specific understanding of each of these two progressions. The high non-crystallin gene expression in the cortex supports the argument that heterogeneity of gene activity may suggest multiple routes to the final functional phenotype, which in the present case is the synthesis of specific crystallins that make the nuclear region of the lens transparent and of high RI.

An important observation is that the correlated crystallin expression starts in the equatorial cells, persists through the highly robust gene activity in the cortical fiber cells (Figure 7B), and finally emerges at the center of the lens in the future visual axis where all other gene activities decline (Figures 7D and S9). We interpret these data to indicate that the highly correlated expression of nine crystallins (out of 17) suggests their refractive role in the future visual axis, whereas the other eight crystallins may have non-crystallin (non-refractive) functions (Bhat, 2004). The expression of crystallins also indicates that gene products of terminal differentiation need not appear at the end but may start at the beginning of this process and may, in fact, have a directive influence on the progression.

Notably, out of the nine crystallins (Figures S2C, 7C, and 7D) expressed in the nuclear fiber cells, five are γ-crystallins, which are known to have high RI increments (0.203 mL/g) (Zhao et al., 2011a) (Mahendiran et al., 2014; Pierscionek et al., 1987). These data provide a tangible and phenotypically rational purpose for the higher expression of these genes in the fiber cells of the lens nucleus/visual axis. The data suggest a deterministic program that leads to the expression of specific crystallin genes in the nuclear region of the developing lens. This relationship between the specific gene activity and the relevant phenotype will need further elucidation.

Although it is rather obvious to focus on the refractive increments of the γ-crystallins, which contribute to transparency, the physiology and structure of the fiber cells as modulated by gene expression in the early lens may have important functions, which contribute to the development of emmetropia (which is critical for the realization of focused functional vision in infancy) (Iribarren, 2015).

Finally, our data point to the existence of a causal continuum that commences with gene activity, which contains differentiation-specific genes, whose expression is sustained from the beginning until the end. The molecular progression detailed here in single fiber cells suggests that the process of differentiation in the ocular lens starts and ends with similar transcriptional profiles; it is highly significant that the two are connected via highly heterogeneous gene activity, which may appear "stochastic" and yet culminate in phenotypically relevant expression. Because we know the spatial location of this heterogeneity, a more

incisive examination may explain how highly variable gene activity may facilitate deterministic emergence of a phenotype in terminal differentiation (Figure 7E).

Limitations of the Study

Procedures used in this investigation, allow us to place an isolated fiber cell within a broad but contiguous region(s) of morphological and molecular activity within the developing lens; the resolution is thus coarse. An analysis of fiber cells with spatial information that places them next to each other would ultimately be required to reveal the finer nuances of the gene activity and the phenotypic organization that makes this tissue functional (transparent). The data presented in this manuscript, however, clearly suggest that the ocular lens makes a meaningful paradigm for the study of the relevance of single cells to an understanding of the emergence of the tissue phenotype; nonetheless, there are vital limitations in these investigations. For example, first, the procedure we have employed to isolate cells is manual and time consuming; there is a need for a method that isolates more cells quicker, in less time, and in a way that maintains the physiological and morphological integrity of individual fibers while keeping their spatial information intact. Second, the use of proteases and/or storage in less-than-optimal buffers (physiological or otherwise) can introduce both visible (morphological) and invisible (molecular) changes, which may not reflect the native state of the cell within the tissue. Third, technical bottlenecks including the size and shape of cells within a tissue at different ages, their RNA content, and library preparations across hundreds and thousands of single cells needs to be standardized for each tissue (Oldham and Kreitzer, 2018) and here in the ocular lens for each developmental stage/ age. Fourth, single-cell transcriptomics has a critical limitation in detecting molecular indicators of cellular interactions, which may only materialize when the cells are in contact with each other, in a fashion that is constrained by the tissue morphology and physiology. This information may only come from the deep sequencing of the whole-tissue RNA.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND SOFTWARE AVAILABILITY

The dataset presented in this manuscript comes from five chips (96x96) IFCs (Biomark, Fluidigm). The raw Cts obtained from the qRT-PCR of single fiber cell RNA is presented here: https://data.mendeley.com/ datasets/mb59r995sk/draft?a=03419801-0a2b-433e-bffdca4bd0a516cd.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Transparent Methods, nine figures, and one table and can be found with this article online at https://doi.org/10.1016/j.isci.2018.11.024.

AUTHOR CONTRIBUTIONS

Conceptualization, S.P.B. and R.K.G.; Methodology, R.K.G.; Investigation, R.K.G.; Writing – Original Draft, S.P.B.; Writing – Review & Editing, R.K.G. and S.P.B.; Validation and Formal Analysis, R.K.G., D.K., X.Z., and D.E.; Resources, S.P.B.; Data Curation, R.K.G. and D.K.; Visualization, R.K.G. and R.K.K.; Supervision, S.P.B.; Project Administration, S.P.B. and R.K.G.; Funding Acquisition, S.P.B.

DECLARATION OF INTERESTS

Authors declare no competing interests.

iScience

CellPress

REFERENCES

Andley, U.P. (2007). Crystallins in the eye: function and pathology. Prog. Retin. Eye Res. *26*, 78–98.

Augusteyn, R.C. (2007). Growth of the human eye lens. Mol. Vis. *13*, 252–257.

Bassnett, S., Shi, Y., and Vrensen, G.F. (2011). Biological glass: structural determinants of eye lens transparency. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 1250–1264.

Bennett, T.M., Zhou, Y., and Shiels, A. (2016). Lens transcriptome profile during cataract development in Mip-null mice. Biochem. Biophys. Res. Commun. *478*, 988–993.

Bhat, S.P. (2003). Crystallins, genes and cataract. Prog. Drug Res. *60*, 205–262.

Bhat, S.P. (2004). Transparency and non-refractive functions of crystallins–a proposal. Exp. Eye Res. *79*, 809–816.

Campbell, M.C., and Hughes, A. (1981). An analytic, gradient index schematic lens and eye for the rat which predicts aberrations for finite pupils. Vision Res. *21*, 1129–1148.

Chakraborty, R., Lacy, K.D., Tan, C.C., Park, H.N., and Pardue, M.T. (2014). Refractive index measurement of the mouse crystalline lens using optical coherence tomography. Exp. Eye Res. *125*, 62–70.

Costello, M.J., Brennan, L.A., Mohamed, A., Gilliland, K.O., Johnsen, S., and Kantorow, M. (2016). Identification and ultrastructural characterization of a novel nuclear degradation complex in differentiating lens fiber cells. PLoS One *11*, e0160785.

Cvekl, A., and Zhang, X. (2017). Signaling and gene regulatory networks in mammalian lens development. Trends Genet. *33*, 677–702.

Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S., Okamoto, J., Johnston, D.M., Qian, D., et al. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat. Biotechnol. *29*, 1120–1127.

De Maria, A., and Bassnett, S. (2015). Birc7: a late fiber gene of the crystalline lens. Invest. Ophthalmol. Vis. Sci. *56*, 4823–4834.

De Maria, A., Shi, Y., Kumar, N.M., and Bassnett, S. (2009). Calpain expression and activity during lens fiber cell differentiation. J. Biol. Chem. *284*, 13542–13550.

Gangalum, R.K., Horwitz, J., Kohan, S.A., and Bhat, S.P. (2012). alphaA-crystallin and alphaB-crystallin reside in separate subcellular compartments in the developing ocular lens. J. Biol. Chem. *287*, 42407–42416.

Glotzbach, J.P., Januszyk, M., Vial, I.N., Wong, V.W., Gelbard, A., Kalisky, T., Thangarajah, H.,

Longaker, M.T., Quake, S.R., Chu, G., et al. (2011). An information theoretic, microfluidic-based single cell analysis permits identification of subpopulations among putatively homogeneous stem cells. PLoS One *6*, e21211.

Goring, D.R., Breitman, M.L., and Tsui, L.C. (1992). Temporal regulation of six crystallin transcripts during mouse lens development. Exp. Eye Res. *54*, 785–795.

Halpern, K.B., Shenhav, R., Matcovitch-Natan, O., Toth, B., Lemze, D., Golan, M., Massasa, E.E., Baydatch, S., Landen, S., Moor, A.E., et al. (2017). Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature *542*, 352–356.

Horwitz, J. (2003). Alpha-crystallin. Exp. Eye Res. *76*, 145–153.

Iribarren, R. (2015). Crystalline lens and refractive development. Prog. Retin. Eye Res. *47*, 86–106.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

Kobayashi, M., Oshima, S., Maeyashiki, C., Nibe, Y., Otsubo, K., Matsuzawa, Y., Nemoto, Y., Nagaishi, T., Okamoto, R., Tsuchiya, K., et al. (2016). The ubiquitin hybrid gene UBA52 regulates ubiquitination of ribosome and sustains embryonic development. Sci. Rep. *6*, 36780.

Lein, E., Borm, L.E., and Linnarsson, S. (2017). The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. Science *358*, 64–69.

Li, J.J., and Biggin, M.D. (2015). Gene expression. Statistics requantitates the central dogma. Science *347*, 1066–1067.

Lieska, N., Krotzer, K., and Yang, H.Y. (1992). A reassessment of protein synthesis by lens nuclear fiber cells. Exp. Eye Res. *54*, 807–811.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell *161*, 1202–1214.

Mahendiran, K., Elie, C., Nebel, J.C., Ryan, A., and Pierscionek, B.K. (2014). Primary sequence contribution to the optical function of the eye lens. Sci. Rep. *4*, 5195.

McAvoy, J.W., and Chamberlain, C.G. (1989). Fibroblast growth factor (FGF) induces different responses in lens epithelial cells depending on its concentration. Development *107*, 221–228.

Nishimoto, S., Kawane, K., Watanabe-Fukunaga, R., Fukuyama, H., Ohsawa, Y., Uchiyama, Y., Hashida, N., Ohguro, N., Tano, Y., Morimoto, T., et al. (2003). Nuclear cataract caused by a lack of DNA degradation in the mouse eye lens. Nature *424*, 1071–1074.

Oldham, M.C., and Kreitzer, A.C. (2018). Sequencing diversity one cell at a time. Cell *174*, 777–779.

Ostrowski, M., Carmo, N.B., Krumeich, S., Fanget, I., Raposo, G., Savina, A., Moita, C.F., Schauer, K., Hume, A.N., Freitas, R.P., et al. (2010). Rab27a and Rab27b control different steps of the exosome secretion pathway. Nat. Cell Biol. *12*, 19–30, sup pp 11–13.

Peetermans, J.A., Foy, B.D., and Tanaka, T. (1987). Accumulation and diffusion of crystallin inside single fiber cells in intact chicken embryo lenses. Proc. Natl. Acad. Sci. U S A *84*, 1727–1730.

Philipson, B. (1969). Distribution of protein within the normal rat lens. Invest. Ophthalmol. *8*, 258–270.

Pierscionek, B., Smith, G., and Augusteyn, R.C. (1987). The refractive increments of bovine alpha-, beta-, and gamma-crystallins. Vision Res. *27*, 1539–1541.

Pierscionek, B.K., and Regini, J.W. (2012). The gradient index lens of the eye: an opto-biological synchrony. Prog. Retin. Eye Res. *31*, 332–349.

Rowan, S., Chang, M.L., Reznikov, N., and Taylor, A. (2017). Disassembly of the lens fiber cell nucleus to create a clear lens: the p27 descent. Exp. Eye Res. *156*, 72–78.

Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature *498*, 236–240.

Skelly, D.A., Squiers, G.T., McLellan, M.A., Bolisetty, M.T., Robson, P., Rosenthal, N.A., and Pinto, A.R. (2018). Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. Cell Rep. *22*, 600–610.

Snijder, B., and Pelkmans, L. (2011). Origins of regulated cell-to-cell variability. Nat. Rev. Mol. Cell Biol. *12*, 119–125.

Subcpyznski, W.K., Mainali, L., Raguz, M., and O'Brien, W.J. (2017). Organization of lipids in fiber-cell plasma membranes of the eye lens. Exp. Eye Res. *156*, 79–86.

Symmons, O., and Raj, A. (2016). What's luck got to do with it: single cells, multiple fates, and biological nondeterminism. Mol. Cell *62*, 788–802.

Treton, J.A., Jacquemin, E., Courtois, Y., and Jeanny, J.C. (1991). Differential localization by in situ hybridization of specific crystallin transcripts during mouse lens development. Differentiation *47*, 143–147.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Ueda, Y., Duncan, M.K., and David, L.L. (2002). Lens proteomics: the accumulation of crystallin modifications in the mouse lens

with age. Invest. Ophthalmol. Vis. Sci. *43*, 205–215.

van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. *9*, 2579–2605.

Weng, J., Liang, Q., Mohan, R.R., Li, Q., and Wilson, S.E. (1997). Hepatocyte growth factor, keratinocyte growth factor, and other growth factor-receptor systems in the lens. Invest. Ophthalmol. Vis. Sci. *38*, 1543–1554.

Wucherpfennig, K.W., and Cartwright, A.N. (2016). Genetic screens to study the immune system in cancer. Curr. Opin. Immunol. *41*, 55–61.

Yamanaka, Y.J., Gierahn, T.M., and Love, J.C. (2013). The dynamic lives of T cells: new approaches and themes. Trends Immunol. *34*, 59–66.

Zaiss, M., Hirtreiter, C., Rehli, M., Rehm, A., Kunz-Schughart, L.A., Andreesen, R., and Hennemann, B. (2003). CD84 expression on human hematopoietic progenitor cells. Exp. Hematol. *31*, 798–805.

Zhao, H., Brown, P.H., Magone, M.T., and Schuck, P. (2011a). The molecular refractive function of lens gamma-Crystallins. J. Mol. Biol. *411*, 680–699.

Zhao, H., Brown, P.H., and Schuck, P. (2011b). On the distribution of protein refractive index increments. Biophys. J. *100*, 2309–2317.

Supplemental Information

Spatial Analysis of Single Fiber Cells

of the Developing Ocular Lens Reveals

Regulated Heterogeneity of Gene Expression

Rajendra K. Gangalum, Dongjae Kim, Raj K. Kashyap, Serghei Mangul, Xinkai Zhou, David Elashoff, and Suraj P. Bhat

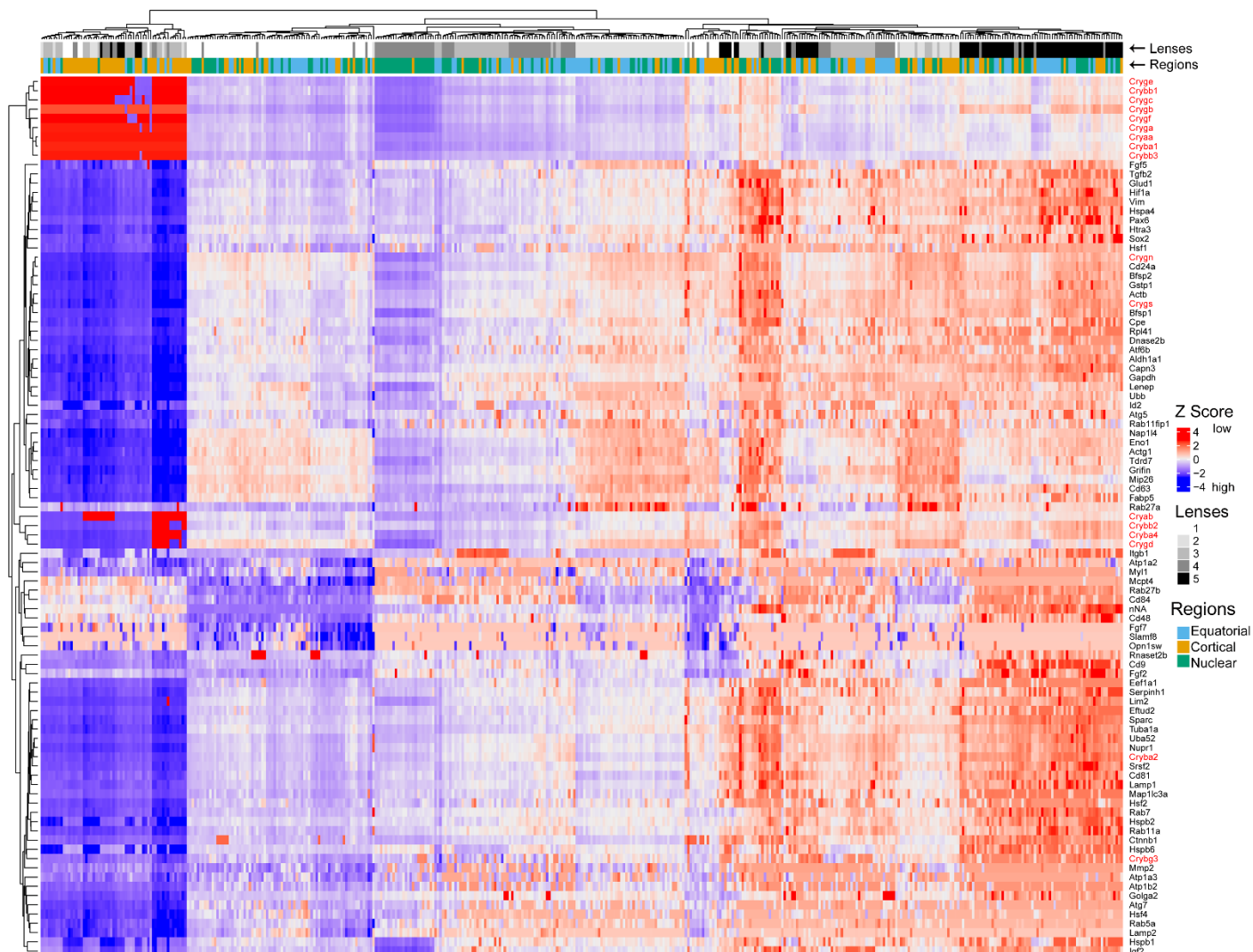**Supplemental Information**

Supplemental Figures and Legends

Figure S1

Figure S1. Complete linkage hierarchical clustering shows heterogeneity of gene expression in single fiber cells, Related to Figure 2.
The Ct value for each gene was converted to a Z score and represented in the heat map (red = low expression and blue-high expression). By using linear mixed effect model, we found that the gene expression in single fiber cells isolated from different regions (Equatorial, Cortical and Nuclear), from each lens correlates well with 95% confidence intervals. The most striking feature of this data is the presence of low crystallin and high non-crystallin expressing cortical fiber cells contributed by all lenses.
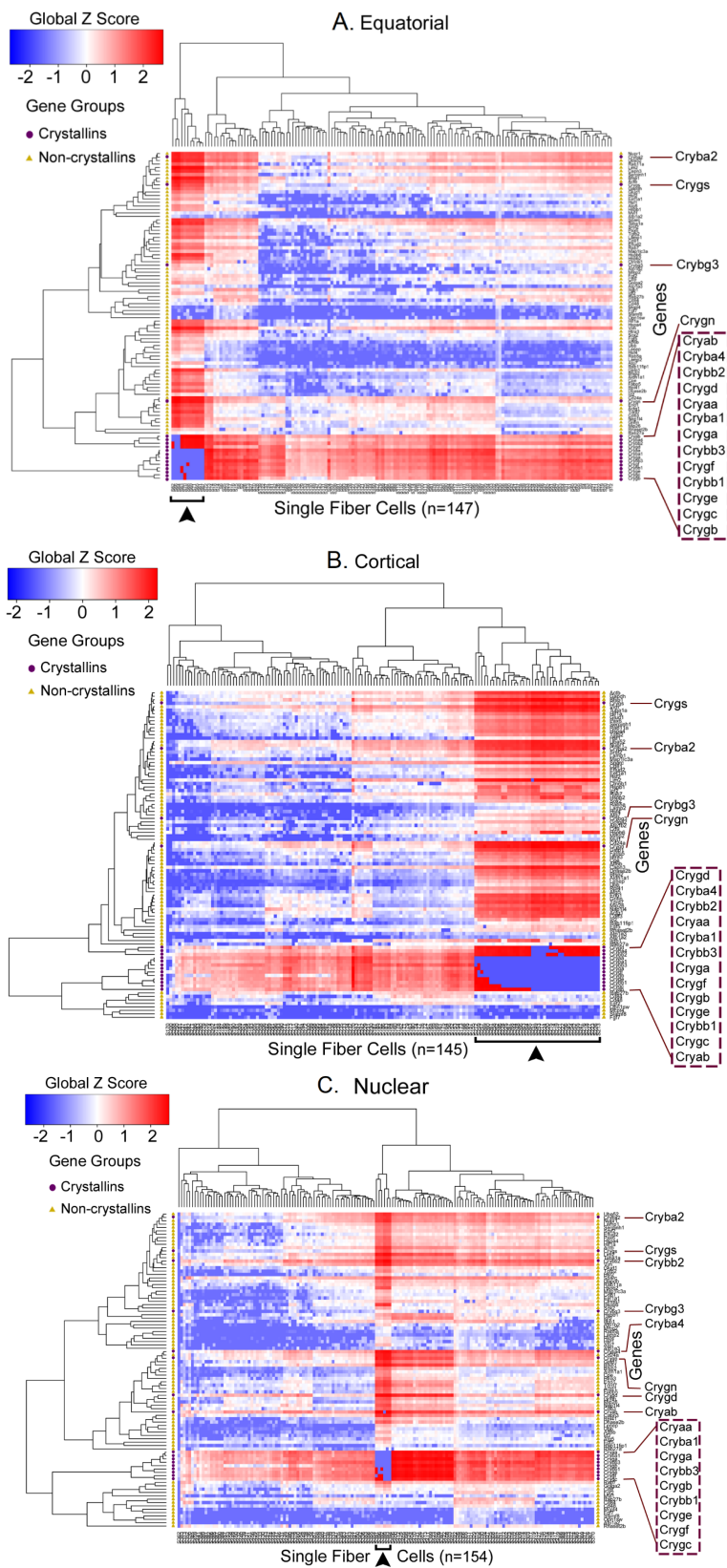
Figures S2 (A-C)

Figure S2 A-C. Unsupervised clustering of gene expression data in single fiber cells isolated from three different regions of the PND02 lens, Related to Figure 2.

(A) Heat map of equatorial fibers (n= 147). Of these fiber cells 92.5% (n=136) express 13 crystallin genes (Cryab, Cryba4, Crybb2, Crygd, Cryaa, Cryba1, Cryga, Crybb3, Crygf, Crybb1, Cryge, Crygc and Crygb (shown in dotted rectangle on the right); note that most of these fiber cells show poor non-crystallin expression patterns except the cells in the extreme left. These cells (n=11), 7.5% of fiber cells in this region, show minimal expression of crystallin genes and high non-crystallin expression (black arrowhead with a bracket); these fibers are part of the 58 fiber cells seen in the collective analysis of 446 fiber cells (Figure 2B, open arrow #2). Sixty three genes (including the 13 crystallins) are expressed in >90% of equatorial fiber cells, whereas 9 genes (Mcpt4, Atp1a3, Hsf4, Atp1a2, Fgf7, Opn1sw, Rab5a, Slamf8 and Lamp2) are only expressed in < 50% of the equatorial fiber cells. (B) Heat map of expression in cortical fiber cells (n= 145). Same 13 crystallins as seen in A above (Figure 1B) are seen here (albeit in a slightly in different order). These cells do not show as high crystallin expression as in A, but 29 % (n=42) of the fiber cells in this region show poor expression of crystallins and high expression of non-crystallins (arrow with a bracket). These cortical fibers are part of the 58 fiber cells (Figure 1B, open arrow #2). (C) Heat map of nuclear fiber cells (n=154). Note that only 9 crystallin genes (Cryaa, Cryba1, Cryga, Crybb3, Crygb, Crygb1, Cryge, Crygf and Crygc) (dotted box on the bottom right) are clustered and expressed in 96% of fiber cells. The other 4% (n=5) that show very low expression of crystallin genes, have high non-crystallin gene expression (black arrowhead with a bracket). These fibers are obviously part of the 58 fiber cells shown in Figure 1B (open arrow #2). In fact, these fiber cells may all be cortical in origin because this is a manual collection and a low percentage of cells may either cross over or be part of a population of cells in transition.

Figure S3

Violin Plot of Gene Expression in Equatorial Fiber Cells

Box Plot of Gene Expression in Equatorial Fiber Cells
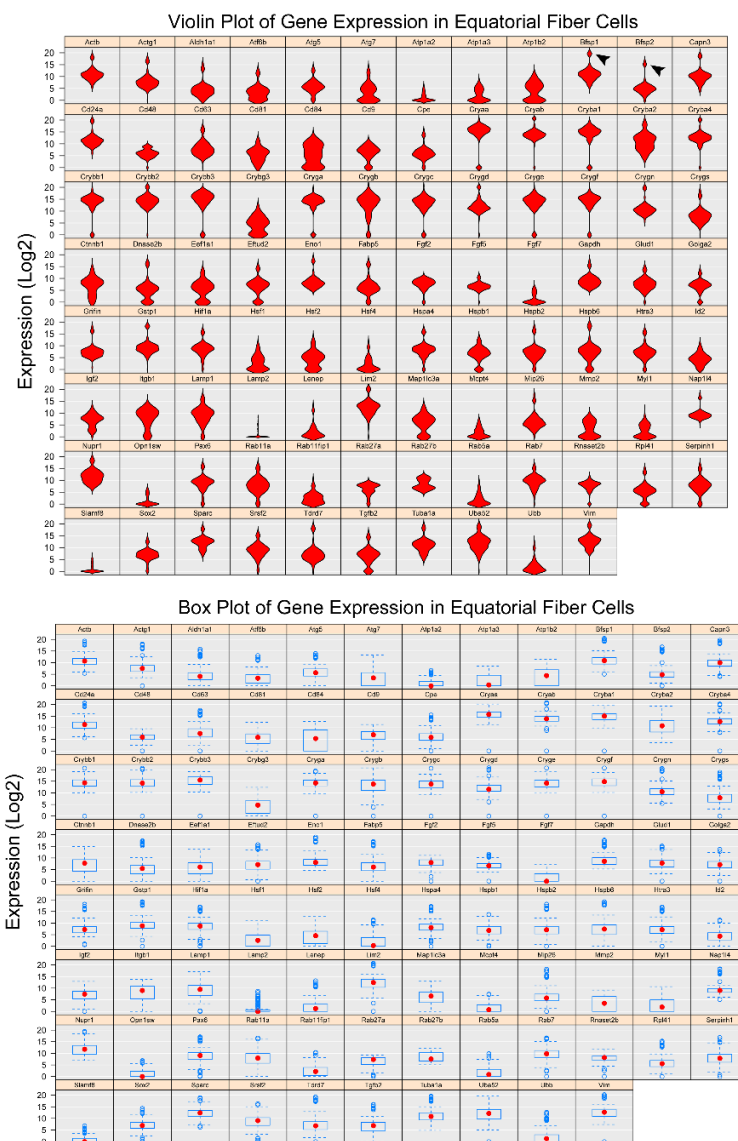
Figure S3. Violin and box plots for the fibers in the equatorial region, Related to Figure 3.

Most crystallins show unimodal expression. Crybg3 and Cryba2 are bimodal and CD84 is robust and multimodal. Most of the non-crystallin gene expression patterns are unimodal and some genes such as Atf6b, Atg7, Atp1b2, Dnase2b, Eef1a1, Eftud2, Hsf1, Hsf2, Rab27b and Map1lc3a, show bimodal expression patterns, suggesting that they are expressed in more than one population of fiber cells. A number of genes show a small distribution blip of high expression distribution (indicated by a black arrowhead in Bfsp1 and Bfsp2). The distributions seen in the equatorial fiber cells are similar to nuclear region fibers (see Figure S5). These distributions are different than those seen in the cortical region fibers (see Figure S4). Equatorial fiber cells show substantially low expression of Atp1a2, Fgf7, Lamp2, Opn1sw, Slamf8 and Ubb levels. Box plots of gene expression in equatorial fiber cells show narrow distribution of gene activity in this region.

Figure S4

Violin Plot of Gene Expression in Cortical Fiber Cells

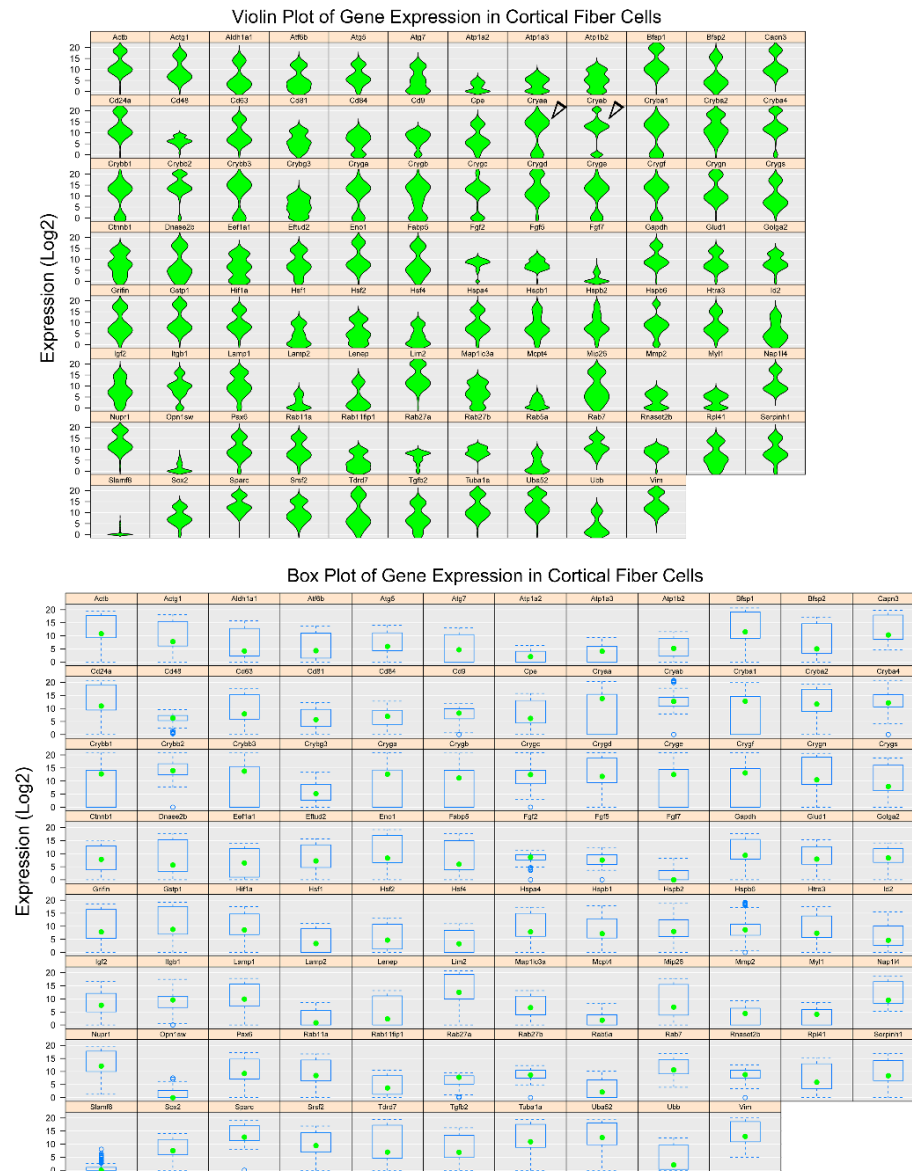Box Plot of Gene Expression in Cortical Fiber Cells

Figure S4. Violin and box plots for the cortical fiber cells, Related to Figure 3.

Most of the cortical fibers (~95%) show bimodal distributions of crystallin as well as non-crystallin genes indicating presence of more than one sub-population of cells with specific expression patterns. These sub-populations of cells exhibit low to high expression log2 values (see corresponding box plots). About 5% of the genes (e.g., Fgf2, Fgf7, Lamp2, Slamf8 and Opn1sw) demonstrate significantly low and unimodal expression profiles. The box plots show broad range of distribution of gene expression in the fiber cells of this region (compare with Figure S3 (equatorial region) and Figure S5 (nuclear region, see below). Cryaa shows a bimodal distribution, while Cryab shows at least three different distributions (violin plot, 2nd row, open arrowheads), corroborating previously suggested independent existence of these two heat shock proteins (Gangalum

et al., 2012). Importantly, multimodality does not distinguish between the terminal differentiation products and other the genes. This is significant because crystallins as well as non-crystallin are mostly unimodal both in the equatorial as well as nuclear regions indicating generally increased heterogeneity in the cortical fiber cell population.

Figure S5

Violin Plot of Gene Expression in Nuclear Fiber Cells

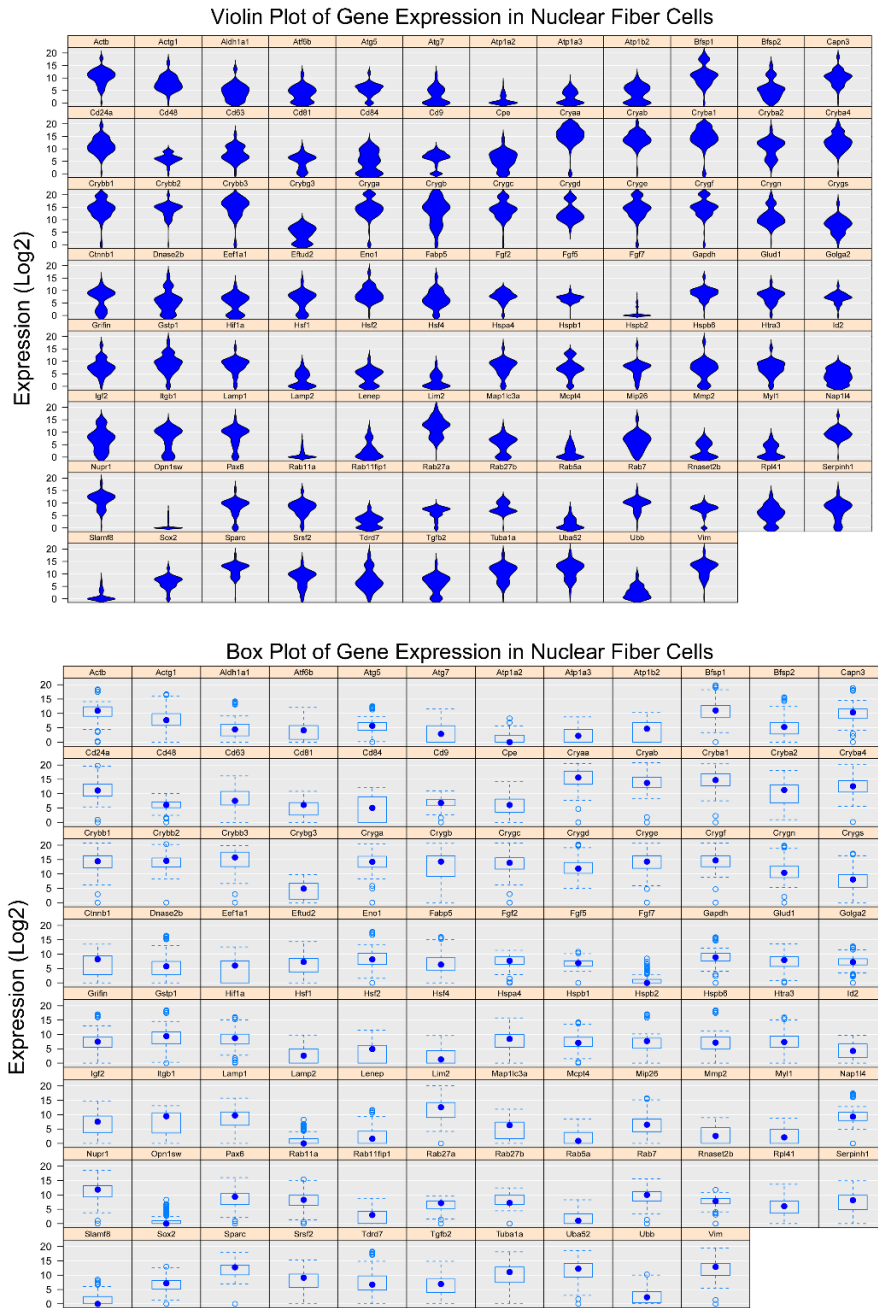Box Plot of Gene Expression in Nuclear Fiber Cells

Figure S5. Violin and box plots for the fiber cells derived from the nuclear region, Related to Figure 3.
The gene expression distributions resemble those seen in the equatorial fiber cells (Figure S3). Atp1a2, Fgf7, Lamp2, Mcpt4, Opn1sw and Slamf8 show very low expression levels. Just as we see in the equatorial fibers, nuclear fiber cells also display narrow range of gene expression distributions. The two regions show much lower heterogeneity than the cortical fiber cells.
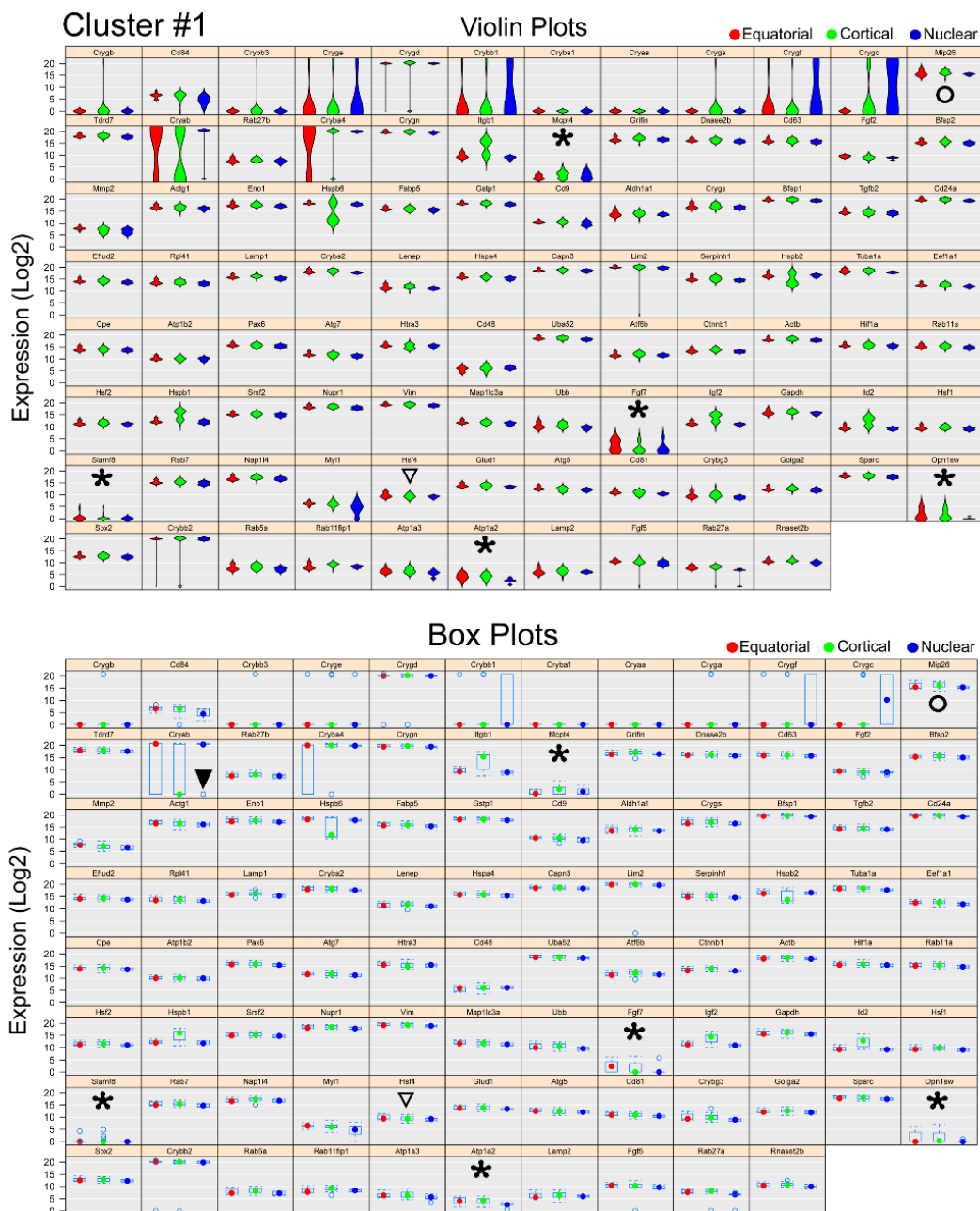
Figure S6

Figure S6. Violin and box plots for the PCA data (cluster #1), presented in Figure 4.

Heat maps and partial violin plots for first three clusters are presented in Figure 2, D-I. For clarity, in the Figure 4 (main text) the data was presented for a set of only 27 genes (17 crystallins + 10 non-crystallins) from each cluster. This cluster contains 58 fiber cells, predominantly from the cortical region; crystallin expression is very low, while the expression of non-crystallin genes is significantly higher (the exceptions are Mctp4, Slamf8, Atp1a2, Fgf7 and Opn1sw (asterisks). Examination of the box plots shows that most non-crystallin present compact median expression. Cryab presents an example of

varied distributions; in the equatorial fibers it shows a very high median (red dot), in the cortical fibers the median is at the lowest level (green dot). Interestingly, in the nuclear fibers, the median expression is high. But this cluster has very few (six) fibers from the nuclear region. One of these fibers shows very low expression (blue circle in the Cryab box plot, ▼arrowhead). Compare Cryab expression with Hsf4 expression, it is very tight (▽). Note also the expression of MIP26, it is high and varies minimally **(O**, top right). This gene is predominantly expressed in these fibers. Red = fibers from equator, green= fibers from the cortical region, blue = fibers from the nuclear region.
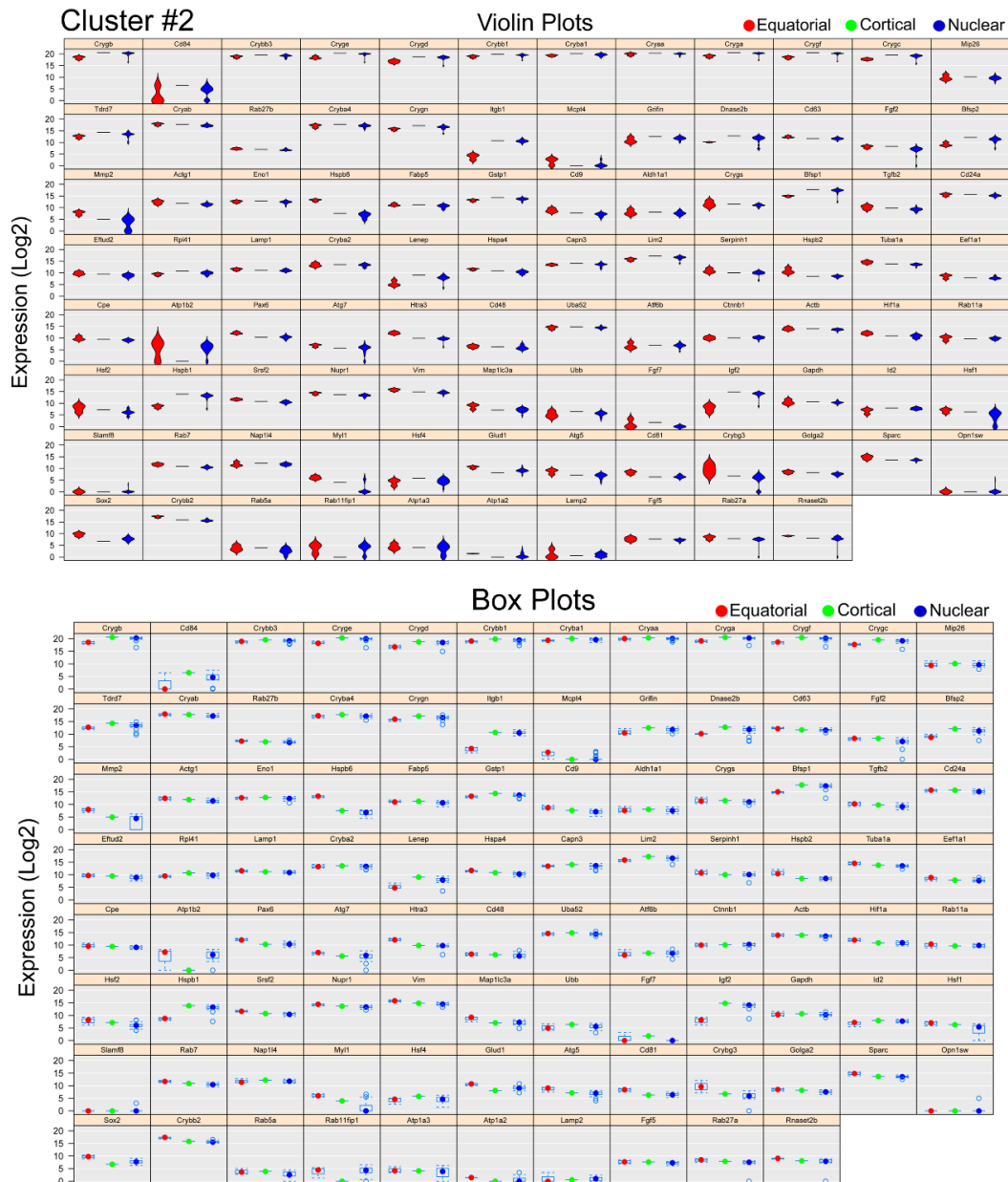
Figure S7

Figure S7. Violin and box plots for the PCA data (Cluster #2) presented in Figure 4.
Cluster #2 contains 28 fiber cells, which are predominantly from the nuclear region. Here, in comparison to the cluster #1 the crystallin expression is high. Note that there are very few cortical (green) fibers in this cluster. The box plot shows very narrow variations in expression.
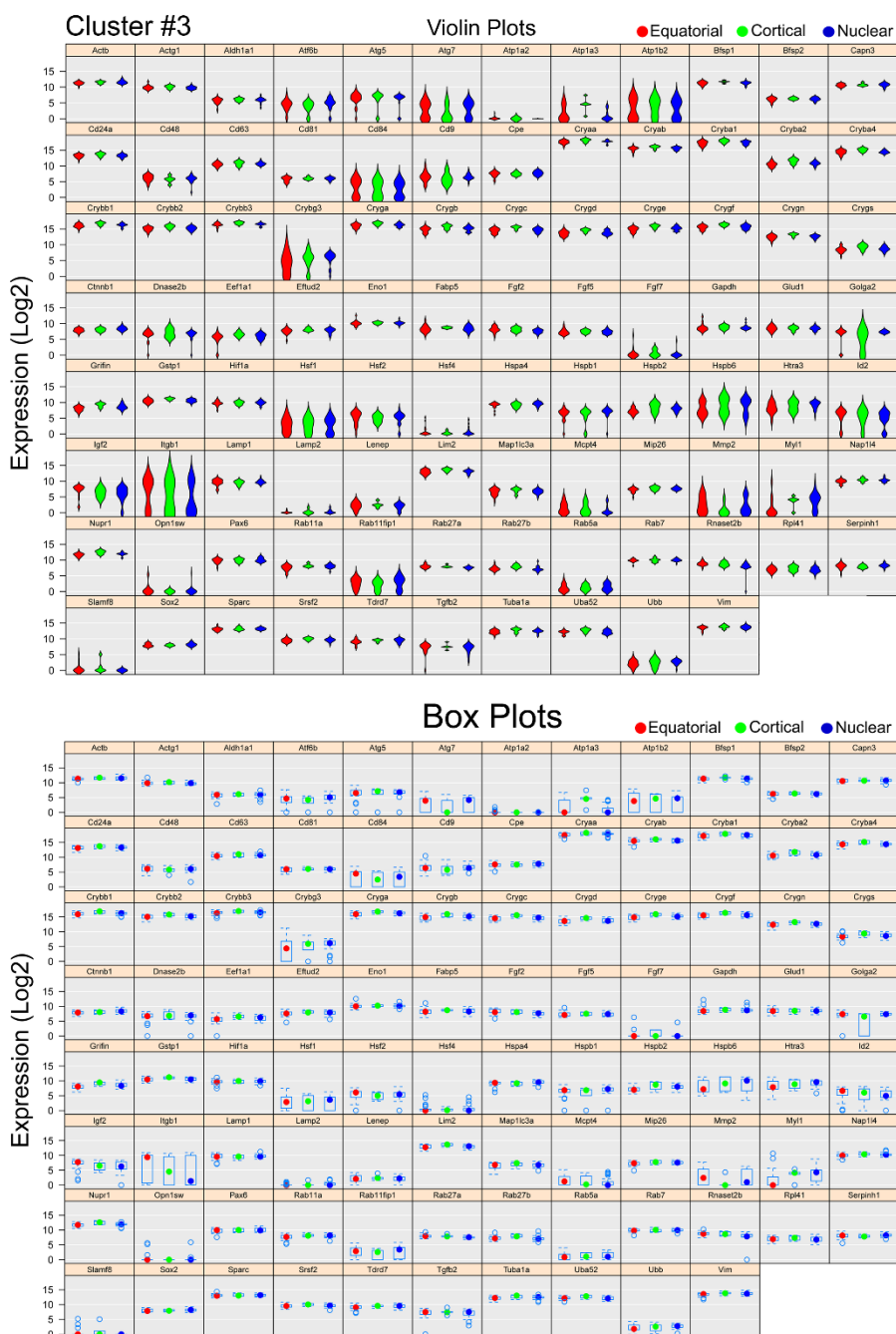
Figure S8

Figure S8. Violin and box plots for the PCA data (cluster #3) presented in Figure 4.

Cluster # 3 contains 48 fibers. Fiber cells from the equatorial region (red) and the nuclear regions (blue) are equally represented (n=21, each). There are six fibers from the cortical region here (green). While clusters #1 and 2 represent two extremes of crystallin expression (low in #1 and high in #2), the cluster # 3 has a mixed range of expression from low to moderate levels.

Figure S9
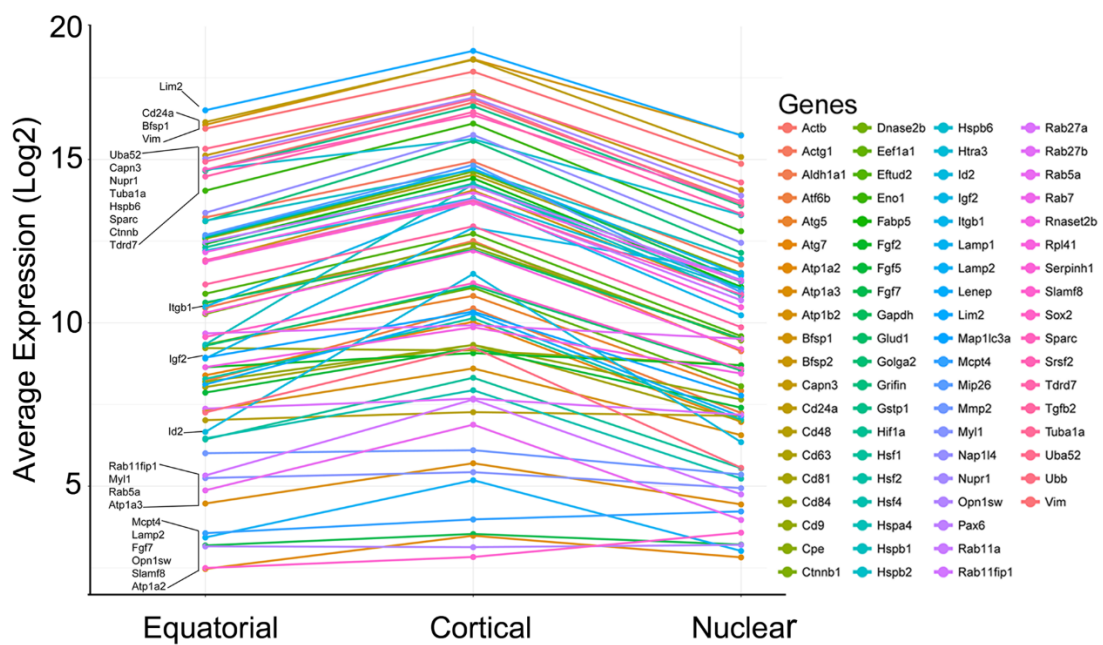
Figure S9. Non-crystallin gene expression decreases in the nuclear fiber cells, Related to Figure 7.
Overall decrease in non-crystallin expression upon transition from the cortical region into nucleus/central lens fibers. Many of these non-crystallin gene activities have been listed on the left hand side (the colors are coordinated with the plot lines). Note that the expression does not go along with the average length of the fibers in different regions (see Figure 1).

Table S1. A list of the genes and gene sequences used for the interrogation of single fiber cells in this study, Related to Figure 2.

Gene Symbol	Forward Primer	Reverse Primer
Actb	CCCTAAGGCCAACCGTGAAA	AGCCTGGATGGCTACGTACA
Actg1	GCTTACACTGCGCTTCTTG	GTGCGGCGATTTCTTCTTC
Aldh1a1	AGACCTGGATAAGGCCATCA	CACTGGGCTGACAACATCA
Atf6b	GTCGACGGAAGATACCTCACA	ACAGGTTTCACTGGAGGAGAC
Atg5	CAGCTCTTCCTTGGAACATCAC	GCATCCTTGGATGGACAGTGTA
Atg7	GCAGTGATGACCGCATGAA	TCGAACCGTGACAGAAAACC
Atp1a2	AGCTGGGCCGAAAATACCAA	TGGGTCCATCTCTAGCCAGAA
Atp1a3	CCCTCAGCAGAAGCTCATCA	ATTCACACCATCGCCAGTCA
Atp1b2	ACTTCTATGCAGGGGCAAAC	CAAAGTGGCCAAGGTTCTCA
Bfsp1	GGAAATGCTGGAACGGCTTA	GAAACTGTGCCTCCAACTGAA
Bfsp2	TCCTGCTGCAGATGGAGAC	CTGAATGGCTGCTCGTTTTCA
Capn3	CCGTCTACAGCACCAGGTTA	CCAGCTTCGTGAAATGGTAGAC
Cd24a	GCTGGGGTTGCTGCTTC	GGAAACGGTGCAACAGATGT
Cd48	GTCTGGTCCTGGAACTGCTAC	TGCCGGTGGTGGCATTTATA
Cd63	CCCTGGGCATTGCTTTTGT	CACTTCGAATACTCTTCACCAGAC
Cd81	GCTGTACCTGGAACTGGGAA	AGCTCCCACAGCAATGAGAA
Cd84	TGGATCTGGTTCCTTTGCCTA	ACTGACTCCCCAAGAATCCC
Cd9	TGTGGAGCTGTACAAGAGTCC	TGGCGAATATCACCAAGAGGAA
Cpe	AAGTGGCAGTTCCTTTTAGCC	CCTCCTCCTTCCTTTCAGAGAA
Cryaa	GGAGATTCACGGCAAACACA	GTCCACATTGGAAGGCAGAC
Cryab	CACGGCAAGCACGAAGAAC	ATCCGGTACTTCCTGTGGAAC
Cryba1	AGCCATGGGTTGGTTCAACA	ATCCAGGGTACTGGTAGCAAAC
Cryba2	TGTGGGTTCCCTGAAAGTCA	ATACTGGTAACCTCGGTAGCC
Cryba4	GAGAGGCTCACCTCCTTCC	AGGAAGTTCTCCTGCTCGAA
Crybb1	ACCGGCTCATGTCCTTCC	TCCATGGTGTTGCCCTTGAA
Crybb2	AGCTCTCTGAGGCCCATCAA	GCCAGTAAAGTTGGGGTTCTCA
Crybb3	AGAAGGTGGGCTCCATCCAA	AACAAACTGCTCCCCACGAAA
Crybg3	CATCGGATCGATTCGTGTCA	AGGAATTGCTGGCCTTTGAA
Cryga	GATGGGTTTCAGCGACTCCA	CCCCGGTAGTCATCTCTCTCA
Crygb	CAGCGACTCCATTCGTTCC	TGAAGTGGAAGCGATCCTGAA
Crygc	CCCACAGAATGCGGCTGTA	TCGCTCAGCTCCATCATGAC
Crygd	CAGTGGATGGGTTTCAGTGAC	CTGGCCTCTGTACTCTTCCC
Cryge	AGGCCAAATGGTGGAGATCA	ATCACGTGGAAGGAGTGGAA
Crygf	CCAGCAGTGGATGGGTTTCA	CTCGCTCGTAGATCCTGATCC
Crygn	AGCTGTGTCAACGCCATCA	GCGGTAGTTTGGCTCCTCATA
Crygs	CGTTGGATGGGCCTTAATGAC	CGTTGAAGTCGCCCTTTTCA
Ctnnb1	CATTGGTGCCCAGGGAGAA	GCCGTATCCACCAGAGTGAAA
Dnase2b	GGTGTCCCTGGATCTGTGAA	CCCTGCGTTCTGTTCCATAC
Eef1a1	GTCGCCTTGGACGTTCTTTT	GCTTTGAATTAGCGGTGGTTTTC

Eftud2	CGGACACCAAAGGGAAGTCTTA	CCAGCTGTGACCTCATCAGAA
Eno1	CGCCTGGCCAAGTACAATCA	CTGAAGGACCTGCCAGCAAA
Fabp5	TTGGTTTACCCAGGATCATTCC	CCTTGAAGAACCCACACACA
Fgf2	TCTTCCTGCGCATCCATCC	GCACACACTCCCTTGATAGACA
Fgf5	ACGTCTCCACCCACTTCCTA	TTCTGGAACAGTGACGGTGAA
Fgf7	GGACCCAGGAGATGAAGAACA	CACCCCTTTGATTGCCACAA
Gapdh	AGACGGCCGCATCTTCTT	TTCACACCGACCTTCACCAT
Glud1	CTGACGTGAGTGTGGATGAA	TTAGCACCTCCAAACGGTAC
Golga2	TCTTGGTGTCACCCCTTCC	CTCCTCCATGAGGTTAGAGCAA
Grifin	AGGAGGAGGTGTCCAGCATA	TGTTCTGCGTCTGCACTCA
Gstp1	TCTACGCAGCACTGAATCC	CTCGAACTGGGAAGTAGACAA
Hif1a	TCGACACAGCCTCGATATGAA	TTCCGGCTCATAACCCATCA
Hsf1	ATGGACTCCAACCTGGACAA	AGGCTCTTGTGGAGACAGAA
Hsf2	GAAACTATTGAGTCCAGGCTTTCA	CTGGGCATGCTTTGCTCTTA
Hsf4	CTGTTTGGGCCACTTCAGAC	CGCTCCCCTCATCTAGCATTA
Hspa4	CTCGGCCTGTACACAGAGAA	AGCATTCTTGGCGTCATTCC
Hspb1	CGGAGATCACCATTCCGGTTA	TGGCTCCAGACTGTTCAGAC
Hspb2	TACTAGTCGCAACAGCAGTCA	TGGGGTTGGCAAATTCGTAC
Hspb6	CCAGTGTGGCGTTACCC	CTCTGGCAAGAAGTGCTTCA
Htra3	ACCATCCAGGACATCGACAA	TGAGTGACCCAGCAGCAA
Id2	ACCCTGAACACGGACATCA	TCGACATAAGCTCAGAAGGGAA
Igf2	GTCTACCTCTCAGGCCGTAC	GGACTGTCTCCAGGTGTCATA
Itgb1	AAGGGCCAACTTGTGAGACA	TGAAGGCTCTGCACTGAACA
Lamp1	GCTTTCAAGGTGGACAGTGAC	CCACAGCAATGGGGATCAAC
Lamp2	ACTACCTGTCTGCTGGCTAC	GTTGTGGCAGGGTTGATGTTA
Lenep	CCTCCAGGTGCCTGTCATTAA	GGATGTAGGCGACTTCCTTCA
Lim2	AACAAGTGCTTCCTGCAGAC	AGGGCAGACAGGATCATGAA
Maf1	TTCAGCACAGCCAGAAGTCA	TGAAAACAGGCTGCAGTTGAC
Map1lc3a	TGAGCGAGTTGGTCAAGATCA	TGGTTGACCAGCAGGAAGAA
Mcpt4	ACAAATCGTTCACCCAAAGTACA	GGACGAGGCAGAGGAATTACA
Mip	GCTGTTGGCTTCTCCCTCA	CATCCCCGCACCAGTGTAATA
Mmp2	CGAGGACTATGACCGGGATA	GGGCACCTTCTGAATTTCCA
Myl1	CCCACCAATGCAGAGGTCAA	AGCTTGCATCATGGGCAGAA
Nap1l4	AAGAAGGCGAGGAAGGTGAA	GGGGTTAACATCAGCATCATCC
Nupr1	AAGGTCGGACCAAGAGAGAA	AACTTGGTCAGCAGCTTCC
Opn1sw	AAGAGCTCCTGTGTCTACAACC	CCTGCACACCATCTCCAGAA
Pax6	GGGGTCTGTACCAACGATAACA	TCTGTTGCTTTTCGCTAGCC
Rab11a	AGGCACAGATATGGGACACA	ATAAGGCACCTACTGCTCCA
Rab11fip1	TGCTCGGTCTCGATAAGTTCC	GGGTTTGGACTTCAGGGTGTA
Rab27a	TGCTTCTGTTCGACCTGACA	ACAGTACGCGTGCATCTGTA
Rab27b	CATCTGCAGCTTTGGGACAC	AAGCCCATGGCATCTCTGAA
Rab5a	CAGGAAACAAAGCTGACTTAGCA	TGATGTCTTAGCTGATGTCTCCA
Rab7	TGGACGACAGACTTGTTACCA	ACCTCTGTAGAAGGCCACAC
Rnaset2b	GGCGTGGCTTAGGGGTTA	ACCTCCACCGACTCCTCA

Rpl41	ACGCCATTAAATAGCAGTAGGC	TCTCATGGCGCAGGAGT
Serpinh1	GATGGGGCACTGCTTGTGAA	TAGGAGCGGGTCACCATGAA
Slamf8	GGTCAAACCTGGACCCAGAC	GCAGCAGTGAACACTTGAACC
Sox2	CCTGCAGTACAACTCCATGAC	TGCGAGTAGGACATGCTGTA
Sparc	GAAACCGTGGTGGAGGAGAC	TGCACCGTCCTCAAATTCTCC
Srsf2	CAAGAGCCCACCCAAGTCT	GTTAAGCCGCTTGCCGATT
Tdrd7	GTTCTGCTCGCTTTCCTTTCA	TTCACAAGGTCAGGGTCATCA
Tgfb2	GCCCATATCTATGGAGTTCAGACA	AGCGGAAGCTTCGGGATTTA
Tuba1a	AGTGTTCGTAGACCTGGAACC	AGTGGCCACGAGCATAGTTA
Uba52	CGAGAATGTCAAGGCCAAGATC	CTGTTTGCCCGCGAATATCA
Ubb	ATTCGGTCTGCATTCCCAGT	AATTGGGGCAAGTGGCTAGA
Vim	GATTTCTCTGCCTCTGCCAAC	CAACCAGAGGAAGTGACTCCA

Transparent Methods

Animals

The mice (C57Blk/6Ncrl) were purchased from Charles River Laboratories International, Inc. Institutional guidelines (Animal Research Committee, University of California, Los Angeles, CA, USA) were strictly followed. Embryonic day 14 pregnant mothers were acclimatized for a week with appropriate diet and water ad libitum. Postnatal two-day (PND02) mice pups were euthanized and the eyes were enucleated in mammalian Ringer's solution (150 mM NaCl, 5.4 mM KC1, 2 mM MgCl2, 2 mM CaCl2, and 20 mM Hepes, pH 7.4, 300 to 310 mOsm). The lenses were processed for staining as described (Gangalum et al., 2014)

Histology

Whole eyes from postnatal day two (PND02) mouse pups were collected in PBS (1X) and fixed overnight at 4°C in 4% paraformaldehyde (made in 1X PBS). The fixed tissue was processed and stained with Hematoxylin and Eosin and imaged with FluoView FV1000 (Olympus Corporation, Tokyo, Japan) (Figure 1A). All images were annotated using Adobe Illustrator (Adobe Creative cloud).

Isolation of Single Fiber Cells from the PND02 mouse ocular Lens

A previously published procedure (Srivastava et al., 1997) for the study of cortical fiber cells was modified. We found that at PND02 the lens fibers come off rather easily upon mild shaking. At PND10 or PND20, it is much harder and takes much longer (3 - 12 hrs), which was not acceptable because we wanted to keep the fiber cells as fresh as possible for RNA extraction.

The lenses with capsules intact were removed from freshly enucleated eyes. A PND02 mouse lens is ~ 900 µm across (about 10-12 mg). The lens was gently rinsed with mammalian Ringer's buffer and then incubated for 10 minutes at 33°C in SHE solution (280 mM Sucrose, 10 mM Hepes, pH 7.4, 10 mM Na-EDTA, 300 to 310 mOsm) containing 0.5 mg/ml trypsin (Life Technologies). The temperature was increased gradually (1°C/min) to 37°C and the incubation continued for next 5 min. At the end of this incubation, the lens with the capsule intact, was rinsed four times with SHE solution (now on without trypsin) and the capsule and lens epithelium were removed gently after making a small incision at the equator. The intact fiber mass was incubated in SHE solution for five minutes without shaking at room temperature. The fiber mass was then transferred to a tabletop shaker with circular rotation (1 revolution per second). The superficial fiber cells detach from the fiber mass. We waited 10 min and collected these fibers manually. These fibers are 'Equatorial' fibers. The fiber mass was quickly rinsed twice with SHE solution and transferred to a new well containing the SHE solution and shaking continued for next 10 minutes to obtain individual fiber cells from the 'Cortical' region. Finally, the remaining fiber mass was gently rinsed once with the SHE solution and a coverslip was placed on top of the fiber mass for next 10 min. The nuclear fibers were dissociated with gentle tapping, on the coverslip and the loose suspension of fibers was transferred to a fresh slide from where the 'Nuclear' region fibers were collected, one at a time. In total we collected 92-96 single fiber cells from each lens (n= 30 to 32 single fiber cells from each region) in about 45 min (Figure 1). We

processed one lens at one time. During this procedure, the fiber cell morphology did not change. The viability of the isolated fiber cells was ascertained by Trypan blue exclusion assay.

RNA isolation and Quantitative PCR (qPCR)

RNA was isolated from single fiber cells (in ~8 to 10 μl of SHE) using single step RNAzol RT method (MRC, Inc. Cincinnati, OH). The entire procedure was performed at room temperature. This method helps isolate pure and intact RNA that is ready for downstream applications without DNase treatment.

Microfluidic qRT-PCR

We used quantitative PCR (qPCR) in the Biomark (Fluidigm Inc., Palo Alto, CA) for assessing the expression for 96 genes (Table S1) in single fiber cell RNAs. Biomark (Fluidigm Inc., Palo Alto, CA). It was important to evaluate if subtle differences in the transcript levels of a gene could be confidently determined without additional validation (routinely done with data obtained with RNA-seq). We initially isolated limited number of individual fiber cells from different regions of multiple lenses and then assayed for the activity of 48 genes including the crystallins. Gene expression obtained with nine different lenses showed remarkable reproducibility (data not shown).

For the data presented in this investigation we used multiple 96 x 96 gene chips and repeated experiments more than twice to assess constancy of the data from one chip to the other. Additionally, we used three dilutions of control whole tissue (lens fiber-mass) RNA in all chips for ascertaining the efficiency for 17 crystallin genes (data not shown)

This method entails eight steps (detailed below) to generate Gene Expression data using 96X96 Integrated Fluidics Circuit (IFC) Dynamic Array chips (Biomark, Fluidigm Inc.). We assembled the reactions using 96 and/or 384 well plates.

Step 1. First strand cDNA was synthesized using 'Transcriptor' universal cDNA master mix (Roche) (reaction volume 8 μl). Thus, a 96 well plate contained 92 individual single fiber cell RNA samples from one lens, 3 dilutions of total PND02 lens RNA and one NTC (no template control).

Step 2. For pre-amplification, 4 μl of cDNA from each sample (Step 1) was individually mixed with 1μl of (10X DELTA gene assay primer mix containing forward and reverse primers (D3 assay design, Fluidigm)) for 95 genes detailed in Table S1. 5 μl of 2X reaction buffer containing platinum *Taq* Mix (ThermoFisher) was added to each reaction making the final reaction volume 10 μl. This reaction was run in the LightCycler 480 (Roche) as follows: 95 °C for 2 min followed by 15 cycles of PCR (95 °C for 15s; 60 °C for 4 min) and a final hold at 4°C (Fluidigm, Advance Development Protocol #B "Fast Gene Expression Analysis Using EvaGreen on the Biomark™ or BioMark HD System). This procedure has been used for all 5 lenses x 96 samples = 480 samples (460 single fiber cells, 15 whole lens total RNA dilutions (3 dilutions on each chip – 1X, 5X and, 10X dilutions) and 5 NTCs (no template controls, one NTC per chip).

Step 3. 5 μl of the PCR product from each sample (step 2) were treated with 2 μl of Exonuclease I mix (1.4 μl of water, 0.2 μl Exonuclease Buffer and 0.4 μl (20 U/μl) Exonuclease, for each reaction) (New England Biolabs Inc., Ipswich, MA). The 96 well plate was vortexed, centrifuged and incubated in a thermal cycler (LightCycler 480,

Roche) at 37 ℃ for 30 min followed by 80 ℃ for 15 min. The reaction was cooled to 4 ℃ until use. After the exonuclease treatment, this product was diluted 20 fold by adding 93 μl of TE (10 mM Tris-HCl, 1.0 mM EDTA, TEKnova, Hollister, CA).

Step 4. Sample pre-mix is prepared (3.0 μl of 2X SsoFast EvaGreen super mix with low ROX dye (Bio-Rad), and 0.3 μl of 20X DNA binding dye sample loading reagent (Fluidigm) per reaction). A specific volume of the sample pre-mix can be prepared based on total number of sample reactions used in an experiment (for example we used 96 reactions). This is added to 2.7 μl of Exonuclease I treated sample (from step 3) for each reaction (total volume thus is 6 μl). The 96-well plate was vortexed for 20 seconds and centrifuged for at 1 min at 1200x g and stored temporarily at -20℃ until use.

Step 5. Assay (primer) mix for each reaction was prepared by mixing 3.0 μl of 2X Assay loading reagent (Fluidigm), 2.7 μl of 1X DNA suspension buffer (TEKnova), and 0.3μl of 100 μM forward and reverse primer mix for each gene, to a total volume of 6 μl. The assay mix was vortexed for 20 seconds and centrifuged for 1 min at 1200x g and stored temporarily at -20℃ until use.

Step 6. A new 96X96 Dynamic Array IFC chip was primed with mineral oil (or control line fluid 150 μl /syringe on both sides). The blue protective film was removed from the bottom of the chip and saved for assay and sample loading indicators. The chip was loaded in the IFC controller HX and then primed (136x script for 96x96) for ~60 min. The chip was removed as soon as it was the priming was done.

Step 7. 5 μl of each assay mix (from step 5) were loaded on the chip, starting from the left side of the chip, column-wise (note the indicator on the top left hand corner of the chip). 5 μl of each sample mix (from step 4) were loaded on the right side of the chip as per the chip template. The chip was placed in the IFC controller and run (the software load mix script (136x) for the 96x96 Dynamic Array IFC) and removed upon completion.

Step 8. The chip was loaded into Biomark HD system using the specific protocol file for gene expression.

Data Analysis

We have merged data from 5 chips (5 lenses x 96 samples=480 samples) using the merge chips command in Fluidigm software and saved one output file in .bml format. The data was then saved as an output file in .csv format and analyzed with SINGuLAR software package (ver 3.6.2), which is compatible with "R" or "R-studio". We have also used other alternative "R" packages (such as corrplot, plotly, ggplot2, calibrate, and reshape2) to generate correlation matrix plots, volcano plots and line plots presented in this manuscript.

Data structure contains three files (1. Data file in .csv file, 2. Sample list in .txt (Equatorial, Cortical and Nuclear fiber cells, total n=446 single fiber cells) and 3. One faulty gene assay (Maf 1) was removed, making total number of genes interrogated =94. Gene list in .txt format (contain two gene groups - 17 Crystallins and 77 Non-crystallins was used to perform "autoAnalysis()" in "R". The autoAnalysis() command saves the data as an object file in .fso format (exp(auto_analysis).fso). This file can be translated from an .fso to .csv format using a macro-enabled converter named "fluidigmSCObjectToExcel", which can be found at the Fluidigm website. The .csv file contains all the information such as method of analysis, number of samples and genes

(assays) used to analyze the data and outlier information as well as Expression (Log2) values. After removing outlier information (such as samples which are above 24 Cts (limit of detection (LoD), or samples that have no amplification). There were 14 fibers with expression log 2 <2. We therefore used 446 fiber cells out of 460 set up for analysis. The linear expression values were calculated as $2^{\wedge(\text{LOD Ct - Measured Ct})}$ or $2^{\wedge\text{Log2Expression}}$. For example, if the LoD is a Ct of 24, then the Expression (log2) of each reaction is calculated to be 24 minus measured Ct value. The singular software (Fluidigm_SC 3.6.2 version) was used to analyze and generate the hierarchically clustered heat maps, violin plots, box plots and for PCA and t-SNE analysis.

The unsupervised hierarchical clustering of heat maps presented in Figure 2 are based on Global Z Scores. The global Z scores were calculated using the formula, $Z = (x − \mu) / (\sigma / \sqrt{n})$, where x= sample mean of a single cell, μ=global mean of all single cells, σ = standard deviation, n=total number of single cells. The clustered heat map of Global Z Score display is based on the sample similarity, which uses normalized expression values with global means and global standard deviations.

The principal component analysis (PCA) algorithm in the SINGuLAR Analysis Toolset uses successive orthogonal transformations (with defined lengths and angles of the vectors) to convert data into a series of principal components (PCs), which explain variance in the data. The variance shown in single-cell samples (2D Scree plot) is represented in 3D PCA plots of the data set. The first two to three PCs explain most of the variation in the data set. Each successive PC in turn explains the next highest variance for the data, under the constraint that its relationship with the previous PC is zero.

Along with the violin plots, at many places we have presented box-whisker plots. These plots in general are used to represent the range of distribution of gene expression. Box-whisker plot is divided into five data regions; the bottom dotted line indicates minimum standard deviation; lower half of the box shows the lower quartile value, this is followed by the median (a filled circle); upper half of the box denotes upper quartile values and the top dotted line shows maximum standard deviation. We have included this data in the supplementary data for analyses of PCA clusters (Figures S6 –S8) and for each region (equatorial, cortical and nuclear fibers) in Figures S3, S4 and S5 respectively.

Correlation analysis shown in this manuscript is the linear association between two variables. Values of the correlation coefficient always lie between -1 and +1 (as Pearson R value), with +1 indicating that two variables are positively correlated and 0 indicates no correlation or linear relationship, while -1 indicates negative correlation (see main text, Figure 7). The correlation coefficient measures only the degree of linear association between two variables but does not necessarily indicate causality. The correlation coefficient between two genes was calculated by the Pearson method (details and instructions on data analysis are provided in the SINGuLAR Analysis Toolset User Guide, Fluidigm Inc)

In Figure S1, the data was processed by removing samples for which more than one third of the genes have expression values below the limit of detection. Based on the limit of detection (LoD, Ct=24), and an outlier identification methodology (samples which display no Ct and samples above 24 Cts were removed from further analysis), we are left with 446 single fiber cells in this data set. To visualize the data, we first standardized the

Ct values for each gene by calculating the gene specific Z-scores. Then we visualized these standardized measurements using a heat map where we color coded mice and regions, and clustered samples and genes using the complete linkage hierarchical clustering method.

To formally test for the region effects, we utilized linear mixed effect models that incorporate fixed region effects (cortical and equatorial vs. nuclear), fixed mouse effects (representing the five mice) and a random mouse-region effect, which accounts for the correlation between samples from the same region of the same mouse. We report the estimates, the 95% confidence intervals, and the p-values of the region effects for each gene. In addition, we used adjusted Type 3 test p-values for comparing the expression levels across all regions. Within each comparison, we corrected for multiple testing using the Benjamini-Hochberg Procedure (Benjamini and Hochberg, 1995) and report the FDR adjusted p-values. We set the FDR threshold at 5%. R software (R version 3.4.1) was used to perform the analyses (Figure S1).

Data and Software Availability

The dataset presented in this manuscript comes from five chips (96x96) IFCs (Biomark, Fluidigm). The raw Cts obtained from the qRT-PCR of single fiber cell RNA is presented here: https://data.mendeley.com/datasets/mb59r995sk/draft?a=03419801-0a2b-433e-bffd-ca4bd0a516cd

RNA-seq

The total RNA from each fiber mass was isolated using RNAzol RT. The integrity and the quality of the RNA was ascertained by using RNA pico series II chip on Agilent 2100 Bioanalyzer. The RNA sample with RNA integrity number (RIN) higher than 9 was used for RNA-seq. The mRNAs were enriched by (poly (A): poly dT) chromatography and fragmented on Covaris M220 Sonicator to an appropriate size (~300 bp). The double stranded cDNA libraries were prepared using Nugen Ovation Ultralow library system DR kit (Cat#0330). The cDNA library was ligated to the adaptors, followed by amplification and sequenced on Illumina HiSeq 2000 with (single read 50 bp) in UCLA Clinical Microarray core facility.

The RNA-Seq reads were mapped to the human hg19 reference (TopHat v. 2.0.12). Reads unmapped to the transcriptome were mapped to the reference genome (ENSEMBL hg19 build). Gene counts were obtained by counting number of sequencing reads overlapping each of the genes (HTSeq v0.6.1).

A plot of the rank correlations between the data obtained with Biomark and total RNA-seq (data not shown) showed good correlation with crystallin abundance between the two, except *Cryba2, Crygs, Crygn* and *Mip26* (which may be because of specific cell-type expression of gene (s)).

***